

## Supplementary Information

### Materials structure-property factorization for identification of synergistic phase interactions in complex solar fuels photoanodes

*Dan Guevarra<sup>1,†</sup>, Lan Zhou<sup>1,†</sup>, Matthias H. Richter<sup>1</sup>, Aniketa Shinde<sup>1,§</sup>, Di Chen<sup>2</sup>, Carla P. Gomes<sup>2</sup>, John M. Gregoire<sup>1,\*</sup>*

<sup>1</sup> *Division of Engineering and Applied Science and Liquid Sunlight Alliance, California Institute of Technology, Pasadena CA, USA.*

<sup>2</sup> *Department of Computer Science, Cornell University, Ithaca, NY, USA.*

<sup>†</sup> *Equal contribution*

<sup>§</sup> *Present address: Nova Measuring Instruments Inc., Fremont, CA, USA.*

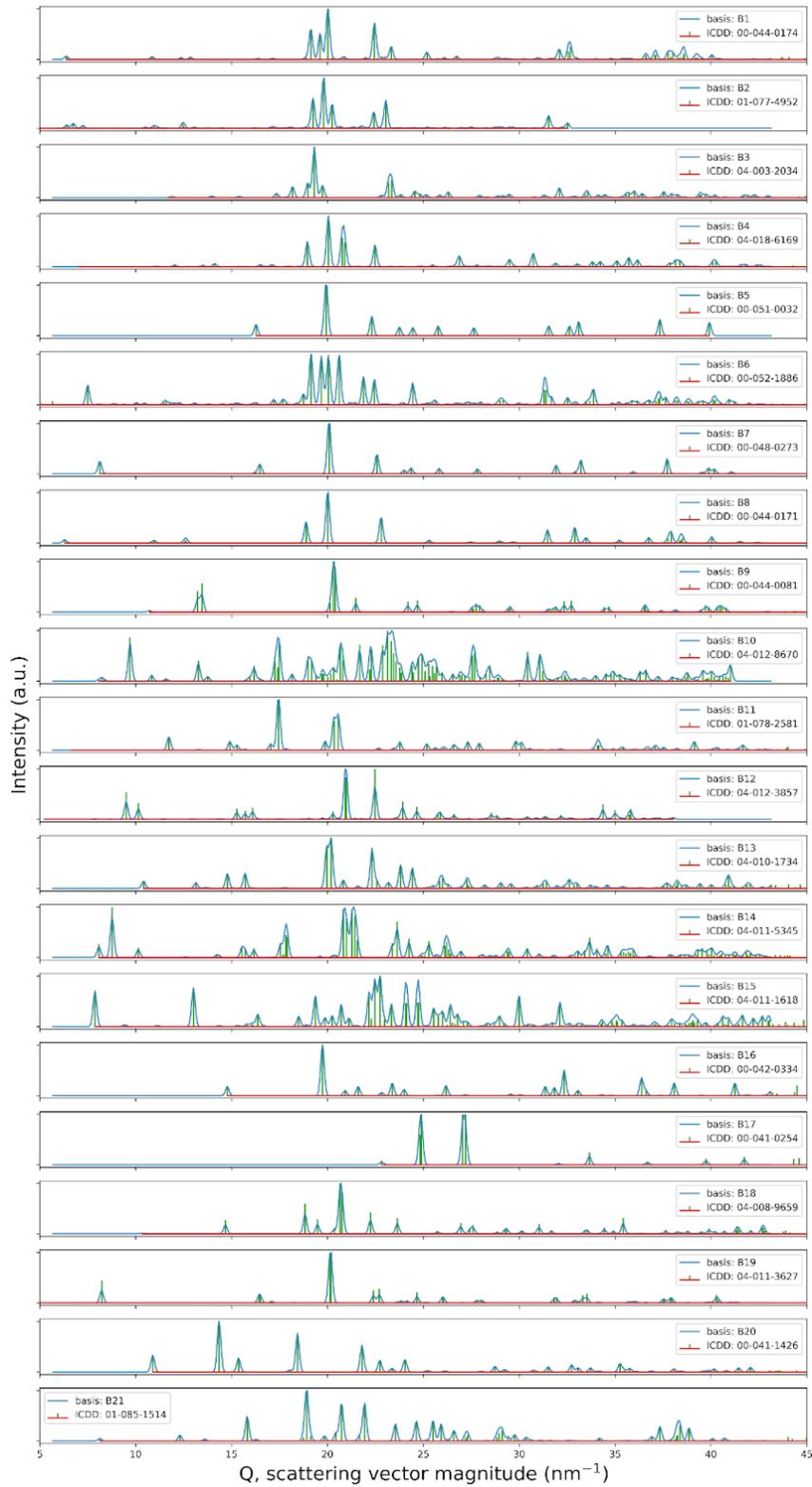
<sup>\*</sup> *Email: gregoire@caltech.edu*

Supplementary Table 1: The 99 candidate phases from the International Crystallography Diffraction Database (ICDD) used for phase mapping, each oxide phase containing Bi, Cu, and/or V.

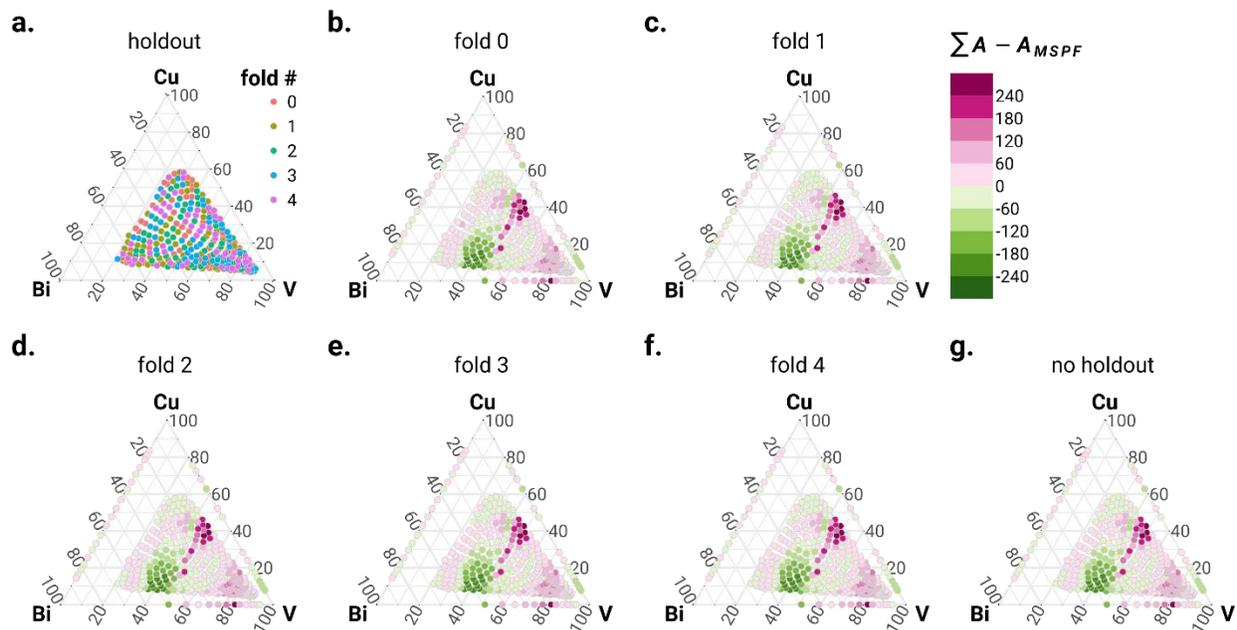
Formula unit	ICDD id	Spacegroup	Formula unit	ICDD id	Spacegroup
Bi(VO <sub>4</sub> )	04-010-5710	I41/a(88)	Cu <sub>2</sub> (V <sub>2</sub> O <sub>7</sub> )	04-014-0715	Fdd2(43)
Bi(VO <sub>4</sub> )	04-010-5713	I2/b(15)	Cu <sub>2</sub> O	01-078-2076	Pn-3m(224)
Bi(VO <sub>4</sub> )	04-015-4206	I41/amd(141)	Cu <sub>2</sub> O	04-020-7578	P42/nm(134)
Bi <sub>1.62</sub> V <sub>8</sub> O <sub>16</sub>	01-082-1957	I4/m(87)	Cu <sub>2</sub> V <sub>2</sub> Bi <sub>0.95</sub> O <sub>8</sub>	04-012-3018	I4/m(87)
Bi <sub>11</sub> VO <sub>19</sub>	00-045-0363	Fm-3m(225)	Cu <sub>2</sub> VBiO <sub>6</sub>	04-012-3857	P21/n(14)
Bi <sub>12</sub> V <sub>2</sub> O <sub>23</sub>	00-044-0174	unkn	Cu <sub>3</sub> (VO <sub>4</sub> ) <sub>2</sub>	04-010-1734	P-1(2)
Bi <sub>14</sub> V <sub>4</sub> O <sub>30</sub>	01-077-4952	C2/c(15)	Cu <sub>3</sub> (VO <sub>4</sub> ) <sub>2</sub>	04-013-2998	P21/c(14)
Bi <sub>14</sub> V <sub>4</sub> O <sub>31</sub>	01-084-5596	P21/c(14)	Cu <sub>3</sub> (VO <sub>4</sub> )	04-016-3668	I-42m(121)
Bi <sub>17</sub> V <sub>3</sub> O <sub>33</sub>	00-052-1476	unkn	Cu <sub>3</sub> V <sub>2</sub> Bi <sub>4</sub> O <sub>14</sub>	04-011-5345	P-1(2)
Bi <sub>23</sub> V <sub>4</sub> O <sub>44.5</sub>	00-047-0733	unkn	Cu <sub>4</sub> V <sub>2</sub> .15O <sub>9.38</sub>	01-070-1696	P212121(19)
Bi <sub>2</sub> O <sub>2.3</sub>	01-076-2477	I4/mmm(139)	Cu <sub>5</sub> V <sub>2</sub> O <sub>10</sub>	04-011-1618	P21/c(14)
Bi <sub>2</sub> O <sub>3</sub>	00-029-0236	P-4b2(117)	Cu <sub>6.5</sub> V <sub>6</sub> O <sub>18.5</sub>	04-017-3174	R-3(148)
Bi <sub>2</sub> O <sub>3</sub>	00-041-1449	P21/c(14)	CuBi <sub>2</sub> O <sub>4</sub>	00-042-0334	P4/ncc(130)
Bi <sub>2</sub> O <sub>3</sub>	00-045-1344	I23(197)	CuO	00-041-0254	C2/c(15)
Bi <sub>2</sub> O <sub>3</sub>	00-050-1088	P-1(2)	CuV <sub>2</sub> O <sub>5</sub>	00-043-0080	C
Bi <sub>2</sub> O <sub>3</sub>	00-058-0356	Pbnb(56)	CuV <sub>2</sub> O <sub>6</sub>	00-027-1136	P21/a(14)
Bi <sub>2</sub> O <sub>3</sub>	00-059-0331	P	CuV <sub>2</sub> O <sub>6</sub>	00-030-0514	C2/m(12)
Bi <sub>2</sub> O <sub>3</sub>	01-080-7655	Pccn(56)	CuV <sub>2</sub> O <sub>6</sub>	00-045-1054	P-1(2)
Bi <sub>2</sub> O <sub>3</sub>	04-003-2034	P21/c(14)	CuV <sub>2</sub> O <sub>6</sub>	04-008-9659	P-1(2)
Bi <sub>2</sub> O <sub>3</sub>	04-015-6853	P-421c(114)	CuVO <sub>3</sub>	00-024-0378	P
Bi <sub>2</sub> O <sub>3</sub>	04-016-7495	P21/c(14)	CuVO <sub>3</sub>	01-071-0860	P1(1)
Bi <sub>2</sub> O <sub>3</sub>	04-018-0027	P63mc(186)	CuVO <sub>3</sub>	04-007-8067	R-3(148)
Bi <sub>2</sub> O <sub>3</sub>	04-018-6169	P21/c(14)	V <sub>0.99</sub> Bi <sub>11.66</sub> O <sub>20</sub>	04-018-2787	I23(197)
Bi <sub>2</sub> VO <sub>5.5</sub>	00-051-0032	I	V <sub>0.90</sub>	04-010-5832	Fm-3m(225)
Bi <sub>2</sub> VO <sub>5.5</sub>	01-088-0870	Aba2(41)	V <sub>2</sub> Bi <sub>4</sub> O <sub>11</sub>	04-011-3627	C2/m(12)
Bi <sub>3.5</sub> V <sub>1.208.25</sub>	00-052-1886	unkn	V <sub>2</sub> Bi <sub>4</sub> O <sub>11</sub>	04-016-4174	A2(5)
Bi <sub>3</sub> V <sub>4.20</sub> 15.5	00-058-0430	I2/a(15)	V <sub>2</sub> O <sub>3</sub>	00-034-0187	R-3c(167)
Bi <sub>4</sub> (V <sub>2</sub> O <sub>11.2</sub> )	01-074-7536	I4/mmm(139)	V <sub>2</sub> O <sub>3</sub>	00-064-0088	Ia-3(206)
Bi <sub>4.1</sub> Cu <sub>0.2</sub> V <sub>1.70</sub> 10.6	00-048-0273	I4/mmm(139)	V <sub>2</sub> O <sub>3</sub>	01-072-5929	I2/a(15)
Bi <sub>4</sub> O <sub>7</sub>	00-047-1058	unkn	V <sub>2</sub> O <sub>5</sub>	00-041-1426	Pmmn(59)
Bi <sub>4</sub> V <sub>2</sub> O <sub>11</sub>	00-044-0357	unkn	V <sub>2</sub> O <sub>5</sub>	00-054-0513	C2/c(15)
Bi <sub>4</sub> V <sub>3</sub> O <sub>12</sub>	00-057-0234	Immm(71)	V <sub>2</sub> O <sub>5</sub>	04-007-2512	Pnma(62)
Bi <sub>6</sub> O <sub>7</sub>	04-015-6628	I4/mmm(139)	V <sub>3</sub> Bi <sub>6</sub> O <sub>16</sub>	04-014-9744	Pnma(62)
Bi <sub>7.38</sub> Cu <sub>0.62</sub> O <sub>11.69</sub>	00-049-1765	P-421c(114)	V <sub>3</sub> O <sub>5</sub>	00-038-1181	P2/c(13)
Bi <sub>7</sub> VO <sub>13</sub>	00-044-0322	I2/m(12)	V <sub>3</sub> O <sub>5</sub>	04-019-7352	Bbmm(63)
Bi <sub>8</sub> V <sub>2</sub> O <sub>17</sub>	00-044-0171	unkn	V <sub>3</sub> O <sub>7</sub>	04-007-0598	C2/c(15)
BiO <sub>1.5</sub>	04-005-4788	Fm-3m(225)	V <sub>4</sub> O <sub>7</sub>	04-005-4524	A-1(2)
BiO <sub>2</sub>	04-006-9431	C2/c(15)	V <sub>4</sub> O <sub>9</sub>	04-007-0884	Pnma(62)
BiV <sub>1.025</sub> O <sub>4+x</sub>	00-044-0081	I2/a(15)	V <sub>5</sub> 1.6O <sub>64</sub>	01-080-3090	I41/amd(141)
BiVO <sub>4</sub>	01-085-1730	Pnca(60)	V <sub>5</sub> O <sub>9</sub>	01-085-1514	B-1(2)
Cu <sub>0.59</sub> V <sub>2</sub> O <sub>5</sub>	04-009-3572	C2/m(12)	V <sub>6</sub> O <sub>13</sub>	01-089-0100	C2/m(12)
Cu <sub>0.64</sub> V <sub>2</sub> O <sub>5</sub>	04-015-4536	Pmmn(59)	V <sub>8</sub> O <sub>15</sub>	01-082-4444	P-1(2)
Cu <sub>1.3</sub> V <sub>9</sub> O <sub>22</sub>	00-046-0362	P21/m(11)	V <sub>9</sub> O <sub>17</sub>	01-082-4445	P-1(2)
Cu <sub>1.5</sub> V <sub>12</sub> O <sub>29</sub>	04-013-7743	C2/m(12)	VBi <sub>2</sub> O <sub>5.5</sub>	04-007-5482	Aea2(41)
Cu <sub>1.82</sub> V <sub>4</sub> O <sub>11</sub>	04-012-3630	Cm(8)	VO <sub>2</sub>	01-076-0675	P42/mnm(136)
Cu <sub>1.98</sub> (V <sub>1.96</sub> O <sub>6.92</sub> )	01-073-4321	C2/c(15)	VO <sub>2</sub>	04-003-2035	P21/c(14)
Cu <sub>11</sub> (VO <sub>4</sub> ) <sub>6</sub> O <sub>2</sub>	04-012-8670	P-1(2)	VO <sub>2</sub>	04-003-4396	C2/m(12)
Cu <sub>2</sub> (V <sub>2</sub> O <sub>7</sub> )	01-078-2581	P1(1)	VO <sub>2</sub>	04-007-0515	C2/m(12)
Cu <sub>2</sub> (V <sub>2</sub> O <sub>7</sub> )	04-011-9703	P-1(2)	VO <sub>2</sub>	04-007-2429	P4/ncc(130)
Cu <sub>2</sub> (V <sub>2</sub> O <sub>7</sub> )	04-012-0620	C2/c(15)			

Supplementary Table 2: The 21 phases identified by DRNets.

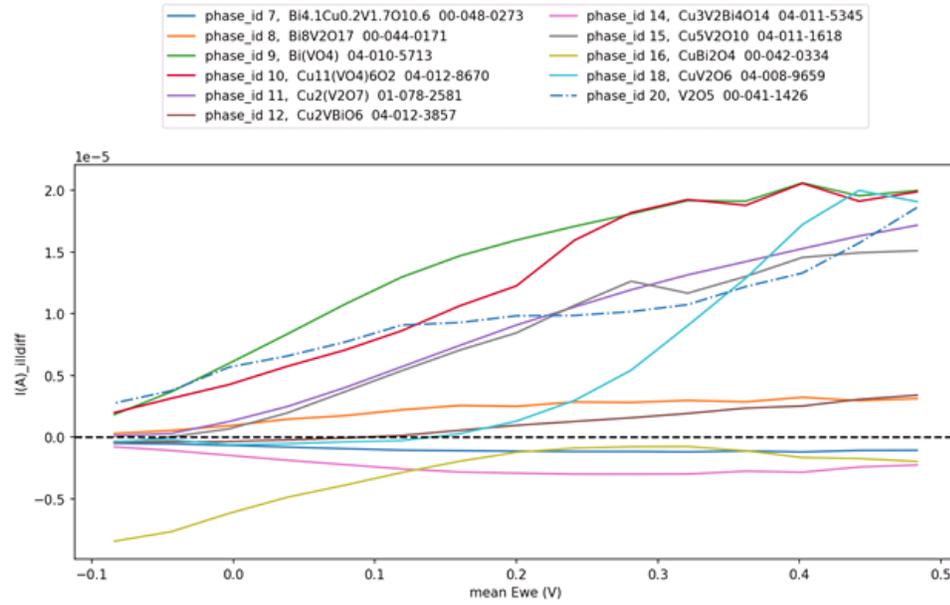
#	Formula Unit	ICDD id	Crystal system	Relative max peak intensity	Cation composition		
					Bi	Cu	V
1	Bi <sub>12</sub> V <sub>2</sub> O <sub>23</sub>	00-044-0174	Monoclinic	801.6	0.86	0.00	0.14
2	Bi <sub>14</sub> V <sub>4</sub> O <sub>30</sub>	01-077-4952	Monoclinic	1326.8	0.78	0.00	0.22
3	Bi <sub>2</sub> O <sub>3</sub>	04-003-2034	Monoclinic	1394.0	1.00	0.00	0.00
4	Bi <sub>2</sub> O <sub>3</sub>	04-018-6169	Monoclinic	1958.3	1.00	0.00	0.00
5	Bi <sub>2</sub> V <sub>0.5</sub> O <sub>5.5</sub>	00-051-0032	Tetragonal	2091.7	0.67	0.00	0.33
6	Bi <sub>3.5</sub> V <sub>1.2</sub> O <sub>8.25</sub>	00-052-1886	Triclinic	345.8	0.74	0.00	0.26
7	Bi <sub>4.1</sub> Cu <sub>0.2</sub> V <sub>1.7</sub> O <sub>10.6</sub>	00-048-0273	Tetragonal	2132.5	0.68	0.03	0.28
8	Bi <sub>8</sub> V <sub>2</sub> O <sub>17</sub>	00-044-0171	Orthorhombic	727.3	0.80	0.00	0.20
9	Bi(VO <sub>4</sub> )	04-010-5713	Monoclinic	863.2	0.50	0.00	0.50
10	Cu <sub>11</sub> (VO <sub>4</sub> ) <sub>6</sub> O <sub>2</sub>	04-012-8670	Triclinic	24.2	0.00	0.65	0.35
11	Cu <sub>2</sub> (V <sub>2</sub> O <sub>7</sub> )	01-078-2581	Triclinic	135.7	0.00	0.50	0.50
12	Cu <sub>2</sub> VBiO <sub>6</sub>	04-012-3857	Monoclinic	259.3	0.25	0.50	0.25
13	Cu <sub>3</sub> (VO <sub>4</sub> ) <sub>2</sub>	04-010-1734	Triclinic	88.9	0.00	0.60	0.40
14	Cu <sub>3</sub> V <sub>2</sub> Bi <sub>4</sub> O <sub>14</sub>	04-011-5345	Triclinic	352.3	0.44	0.33	0.22
15	Cu <sub>5</sub> V <sub>2</sub> O <sub>10</sub>	04-011-1618	Monoclinic	45.5	0.00	0.71	0.29
16	CuBi <sub>2</sub> O <sub>4</sub>	00-042-0334	Tetragonal	64.2	0.67	0.33	0.00
17	CuO	00-041-0254	Monoclinic	177.8	0.00	1.00	0.00
18	CuV <sub>2</sub> O <sub>6</sub>	04-008-9659	Triclinic	83.9	0.00	0.33	0.67
19	V <sub>2</sub> Bi <sub>4</sub> O <sub>11</sub>	04-011-3627	Monoclinic	1115.5	0.67	0.33	0.00
20	V <sub>2</sub> O <sub>5</sub>	00-041-1426	Orthorhombic	101.6	0.00	1.00	0.00
21	V <sub>5</sub> O <sub>9</sub>	01-085-1514	Triclinic	48.9	0.00	1.00	0.00



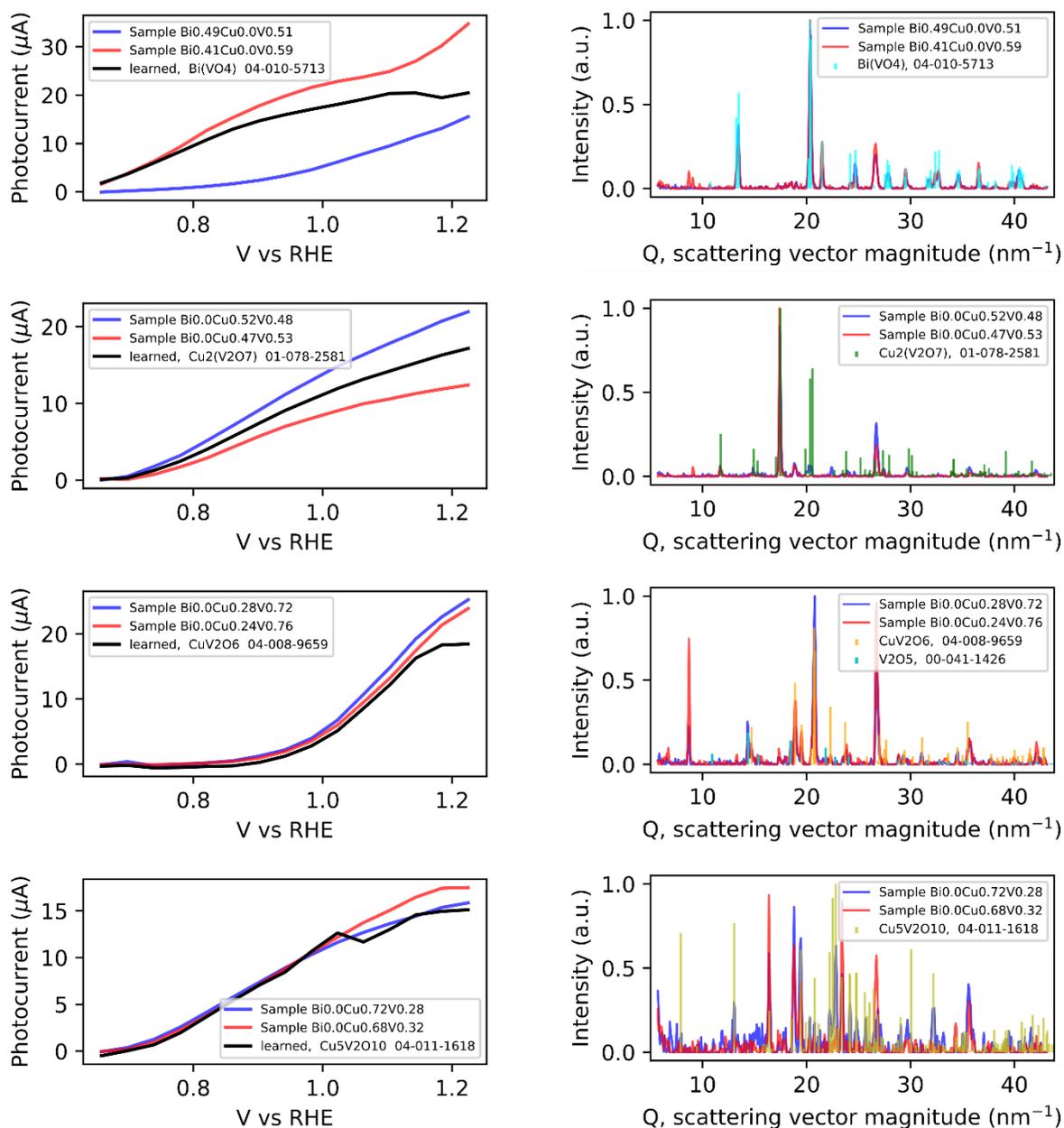
Supplementary Figure 1: The basis patterns learned by DRNets compared to the ICDD stick patterns for each of the identified 21 phases.



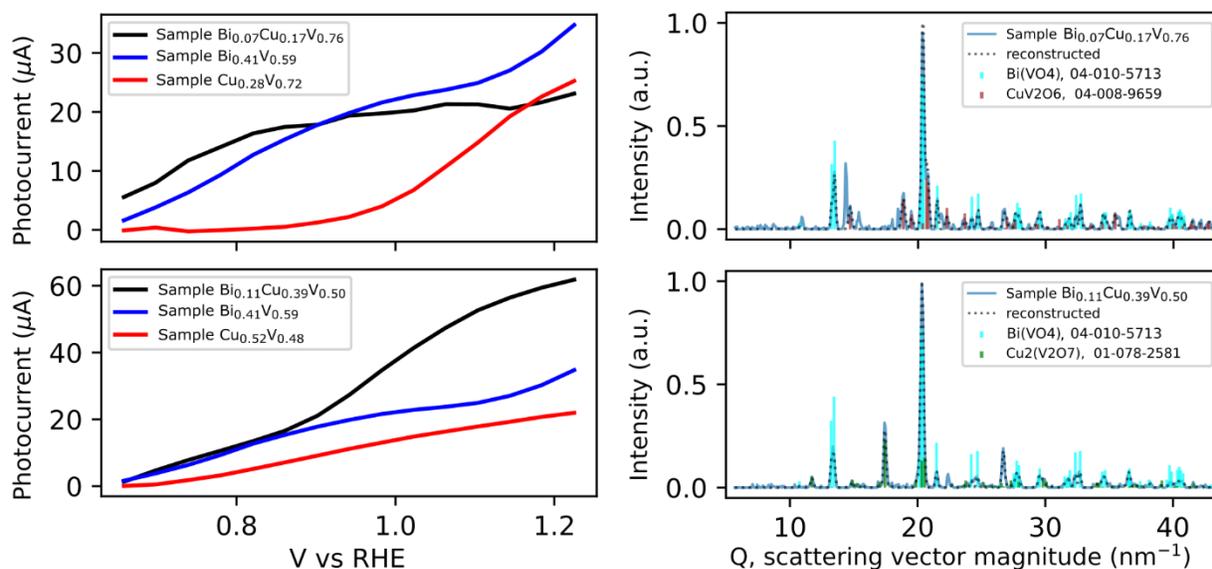
Supplementary Figure 2: Sensitivity analysis of the MSPF results. The 2-cation spaces are relatively sparse samples within approximately 2 representative samples for high-phase-purity samples, as indicated by Supplementary Figure 4. To assess the suitability of the sampling density in the 3-cation space, we repeat the 2<sup>nd</sup> stage matrix factorization using a random 5 folds of the 3-cation data. **a** The 5-fold membership of each 3-cation composition resulting from random selection. **b-f** For each fold, the corresponding 20% of the data points were withheld from the matrix factorization. Using the resulting photoactivity bases, the photoactivity reconstruction and residual analysis was performed in the same manner as for the full dataset. Figure 3 contains the residual from the full-dataset analysis for 3 potential ranges, and for compactness this figure uses the total residual over all potentials, which is shown in **g**. The residual map from each fold is numerically distinct, although the differences are negligible compared to the composition trends in residuals from which the primary conclusions were drawn. This analysis supports the suitability of the sampling of the composition space for MSPF analysis as well as the appropriate construction of the model to avoid overfitting the data.



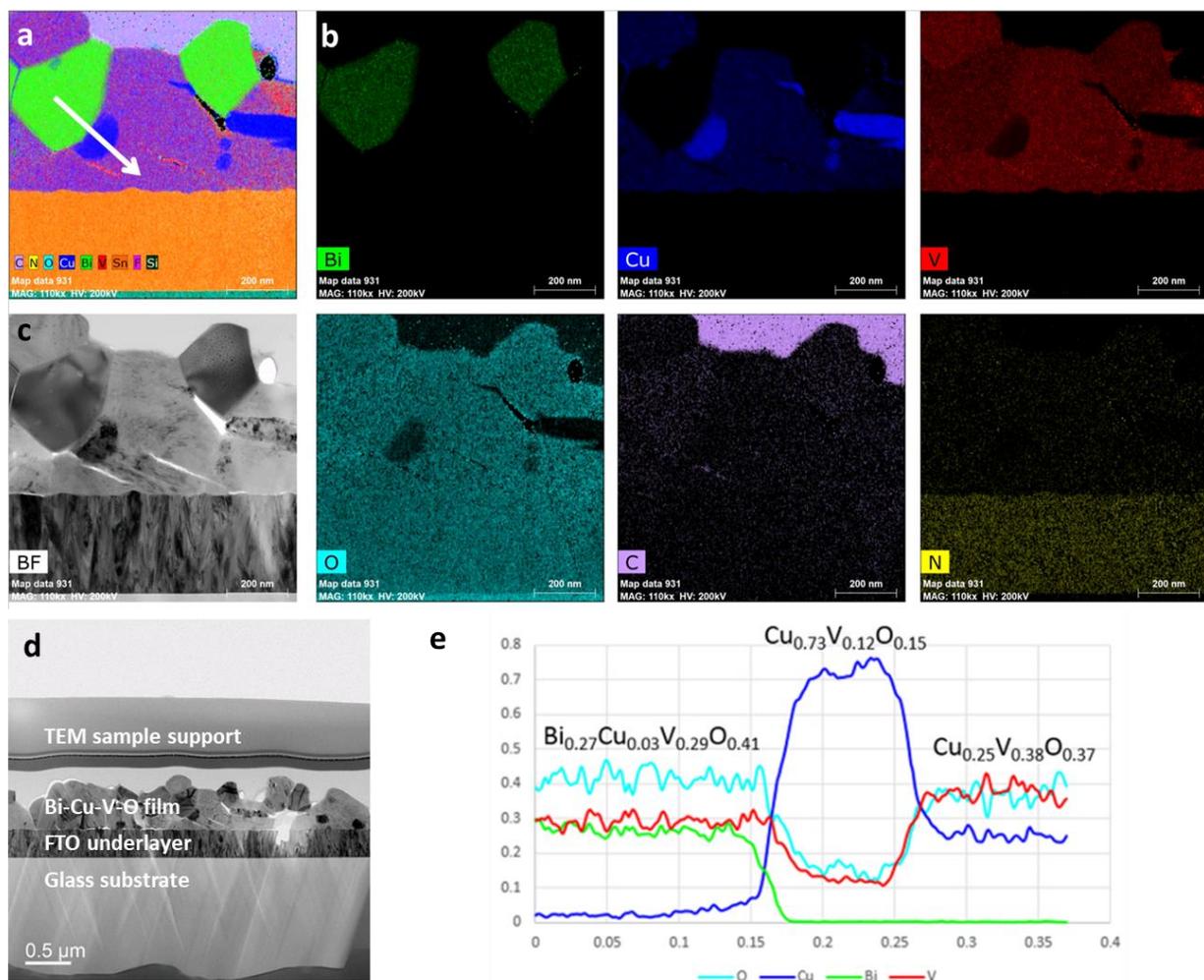
Supplementary Figure 3: The materials property factorization results for each of the 11 phases. Note that these representative patterns are learned from mixed-phase samples, so these data are not necessarily what would be obtained with a given phase-pure material.



Supplementary Figure 4: Among the photoactivity basis components learned via matrix factorization from 2-cation compositions, these are the 4 with appreciable anodic photocurrent. For each of these 4 phases, (left) the PEC basis component is shown with the measured data from 2 photoelectrode samples, which are the 2 samples with phase purity for the respective phase above 80% and with cation composition closest to the formula unit, and (right) the XRD patterns for the 2 samples are shown along with the reference pattern for the respective phase.

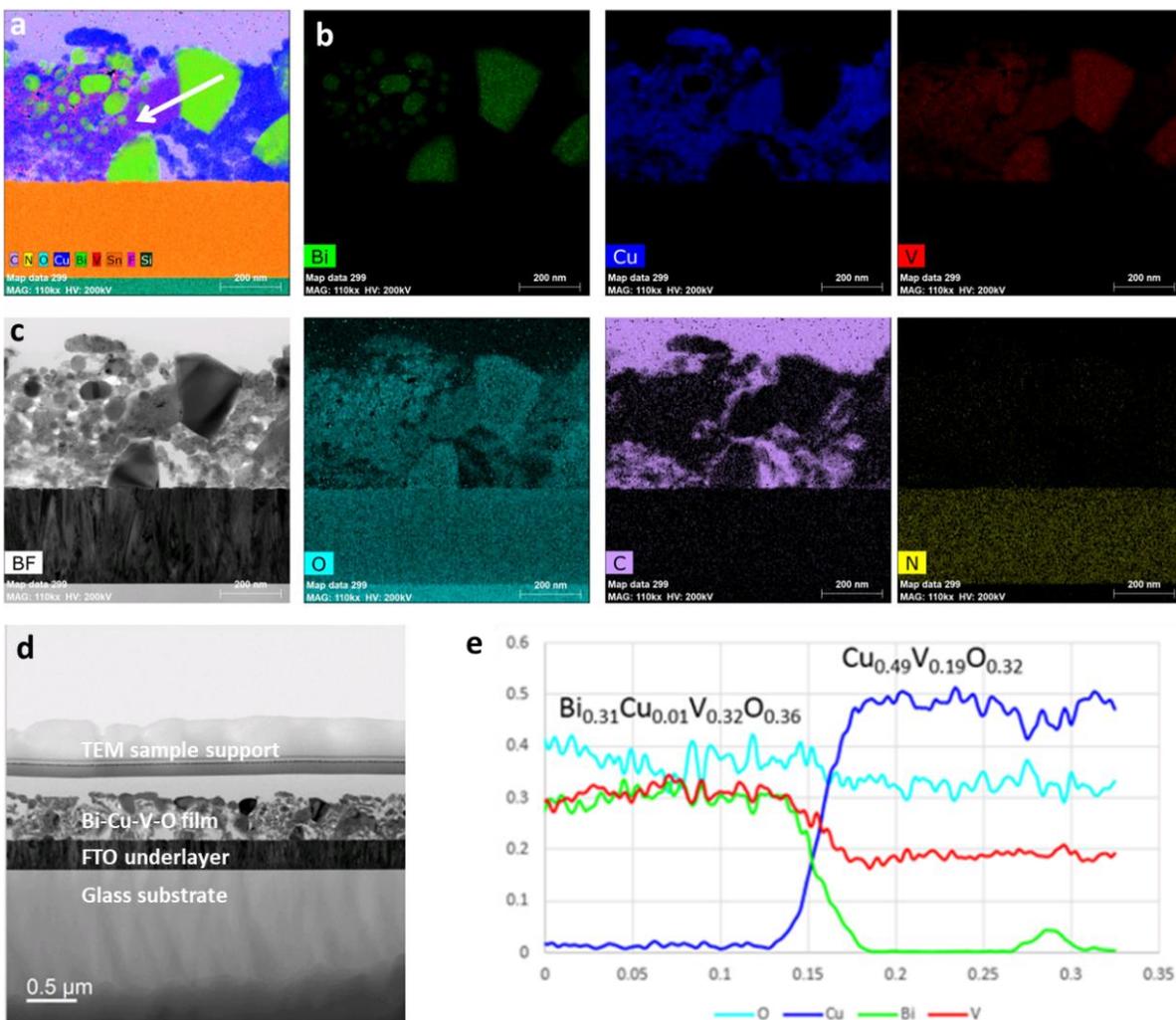


Supplementary Figure 5: For the 2 samples in Figure 4, the measured photoactivity is shown – in this figure only measured photoactivity is shown as opposed to the Figure 4 that additionally contains MSPF decomposition of the photoactivity. Each of these samples contains 2 primary phases, as illustrated by the XRD patterns at right and the DRNets reconstruction of the XRD pattern based on the peaks from the reference patterns from the respective phases. For each sample, the photoactivity plot on the left shows the highest measured photoactivity for each of the 2 containing phases, i.e. for each phase, the photoactivity of the sample with >80% phase purity and highest photoactivity is shown, which is a subset of the data shown in Supplementary Figure 4. This confirms that the photoelectrodes identified through the MSPF residual analysis are indeed substantially more photoactive than the high-phase-purity photoelectrodes of the corresponding phases, where (top) Bi<sub>0.07</sub>Cu<sub>0.17</sub>V<sub>0.76</sub> was selected based on the photoactivity at low applied bias and (bottom) Bi<sub>0.11</sub>Cu<sub>0.39</sub>V<sub>0.50</sub> was selected based on the photoactivity at high applied bias. In each case, the mixed-phase photoelectrode outperforms each pure phase in the respective potential window. In addition, the shape of the photoactivity vs applied bias for each mixed-phase sample is distinct from those of the pure-phase samples, further supporting that an electronic interaction among the 2 phases gives rise to the improved photoactivity.



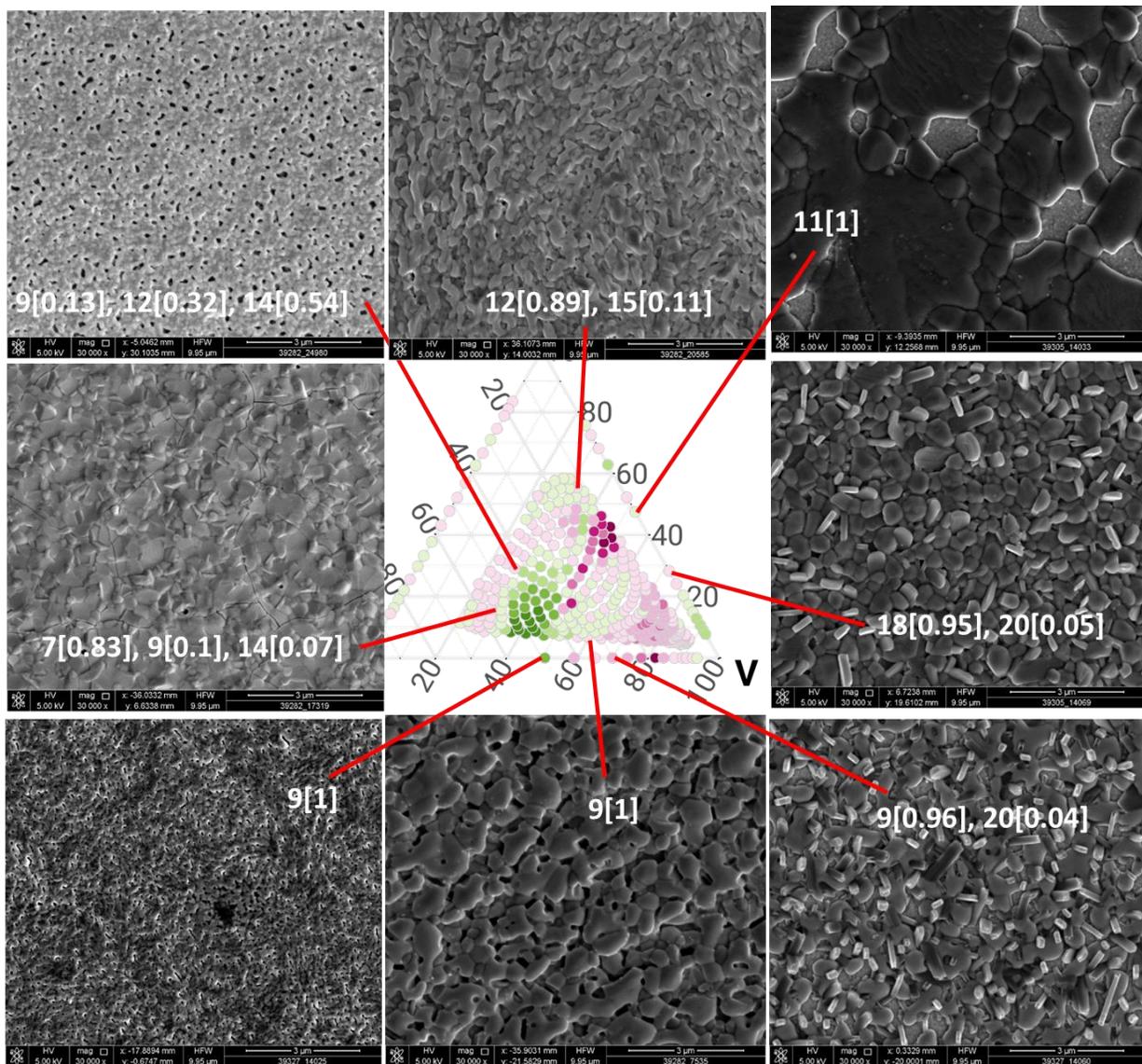
Supplementary Figure 6. A cross section TEM analysis of the  $\text{Bi}_{0.07}\text{Cu}_{0.17}\text{V}_{0.76}$  photoanode after photoactivity measurements. **a** Composition color image of elemental maps to identify different composition domains. **b** Individual color maps for the 6 elements of interest. **c** The brightfield image of the same film region. **d** A lower-magnification bright field image showing the film stack. **e** The cation composition along the white arrow from **a** showing different composition domains that are each labelled by their average composition. The DRNets analysis of the XRD results indicated that the film is composed of 33%  $\text{BiVO}_4$  and 67%  $\text{CuV}_2\text{O}_6$ , which is consistent with the primary composition domain being a V-rich mixture with Cu but little Bi and the secondary composition domain having similar Bi and V concentrations with little Cu. In addition, there are small composition domains that are Cu-rich and have no corresponding analogue in the DRNets results. Supplementary Figure 5 shows some minor XRD peaks that are unexplained by the DRNets solution, illustrating an imperfection in the phase map used for the MSPF analysis. Given that the unexplained composition domain is a small fraction of the film and that all Cu-rich phases have substantially less photoactivity than the 2 primary phases ( $\text{BiVO}_4$  and  $\text{CuV}_2\text{O}_6$ ), this imperfection in the phase map does not alter the conclusions based on the residual analysis. In future work studying the nature of the electronic interactions that give rise to the improved photoactivity, the role of this Cu-rich domain must also be

considered. The high-level summary of the microstructure from the aggregation of our data is that the film thickness varies between 300 and 500 nm with ~200 nm-wide  $\text{BiVO}_4$  domains embedded in a  $\text{CuV}_2\text{O}_6$ -like matrix with small domains of Cu-rich precipitates.



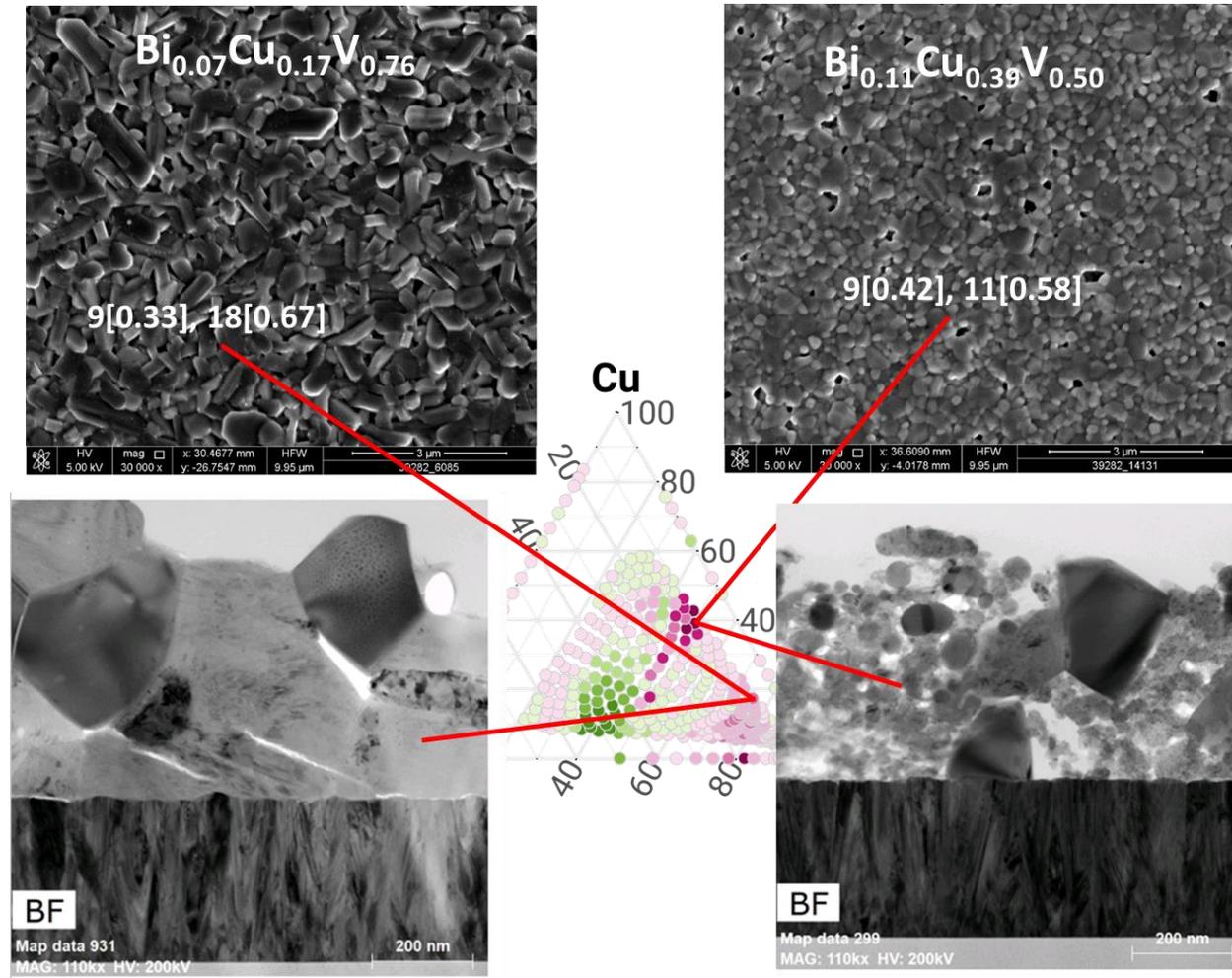
Supplementary Figure 7. A cross section TEM analysis of the  $\text{Bi}_{0.11}\text{Cu}_{0.39}\text{V}_{0.50}$  photoanode after photoactivity measurements. **a** Composition color image of elemental maps to identify different composition domains. **b** Individual color maps for the 6 elements of interest. **c** The brightfield image of the same film region. **d** A lower-magnification bright field image showing the film stack. **e** The cation composition along the white arrow from **a** showing different composition domains that are each labelled by their average composition. The DRNets analysis of the XRD results indicated that the film is composed of 42%  $\text{BiVO}_4$  and 58%  $\text{Cu}_2\text{V}_2\text{O}_7$ , which is consistent with the primary composition domain being a V-rich mixture with Cu but little Bi and the secondary composition domain having similar Bi and V concentrations with little Cu. In addition, there are small composition domains that are Cu-rich and have no corresponding analogue in the DRNets results. Supplementary Figure 5 shows some minor XRD peaks that are unexplained by the DRNets solution, illustrating an imperfection in the phase map used for the

MSPF analysis. Given that the unexplained composition domain is a small fraction of the film and that all Cu-rich phases have substantially less photoactivity than the 2 primary phases ( $\text{BiVO}_4$  and  $\text{Cu}_2\text{V}_2\text{O}_7$ ), this imperfection in the phase map does not alter the conclusions based on the residual analysis. In future work studying the nature of the electronic interactions that give rise to the improved photoactivity, the role of this Cu-rich domain must also be considered. The high-level summary of the microstructure from the aggregation of our data is that the film thickness varies between 300 and 500 nm where  $\text{BiVO}_4$  domains span 10 nm to 200 nm and are embedded in a  $\text{Cu}_2\text{V}_2\text{O}_7$ -like matrix that exhibits variability in both microstructure and Cu:V composition.



Supplementary Figure 8: SEM images of as-synthesized thin films with a constant 3 μm scale bar to observe variations in morphology over a broad range of compositions. The total residual map from Supplementary Figure 2g is shown with red lines indicating the composition corresponding to each image. The images are labelled by the DRNets phase constitution, a list wherein the phase # from Supplementary Table 2 is followed by the molar concentration of the phase in brackets. The bottom 3 images are all high purity  $\text{BiVO}_4$  and show a substantial variation in morphology over a relatively small composition window. The bottom-right image shows a small phase fraction of  $\text{V}_2\text{O}_5$ , which is apparent in the SEM image as the needle-like morphology that is common in  $\text{V}_2\text{O}_5$  thin films. Similar needles are also observed in the middle-right image, which has a small phase fraction of  $\text{V}_2\text{O}_5$  and majority phase  $\text{CuV}_2\text{O}_6$ . The upper-right image is high-phase-purity  $\text{Cu}_2\text{V}_2\text{O}_7$  and shows the largest characteristic grain size as well as some exposure of the underlying FTO. In the other images, sometimes a mixture of phases is

evident in a mixture of characteristic grain morphologies, but for example the upper-left sample containing a mixture of 3 phases appears to have a uniform morphology with appreciable void density.



Supplementary Figure 9. Morphology characterization of the 2 compositions of interest from residual analysis: (left)  $\text{Bi}_{0.07}\text{Cu}_{0.17}\text{V}_{0.76}$  and (right)  $\text{Bi}_{0.11}\text{Cu}_{0.39}\text{V}_{0.50}$ , with format and notation analogous to that of Supplementary Figure 8. The top image for each composition is an SEM image with the same scale as the images from Supplementary Figure 8, and the bottom image for each composition is an enlarged bright field TEM image for the respective composition, Figures (left) 6c and (right) 7c. The combined SEM and TEM characterization shows that the  $\text{Bi}_{0.07}\text{Cu}_{0.17}\text{V}_{0.76}$  sample has substantially larger grains, whereas  $\text{Bi}_{0.11}\text{Cu}_{0.39}\text{V}_{0.50}$  has some large grains with a broad range of smaller grains and substantially higher void density. This range in morphology is comparable to that observed with the high-phase-purity  $\text{BiVO}_4$  samples in Supplementary Figure 8, indicating that morphology alone does not account for the substantial residuals from the MSPF analysis.